(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 November 2006 (30.11.2006)

PCT

(10) International Publication Number
WO 2006/127040 A2

(51) International Patent Classification:
C40B 40/10 (2006.01)

(21) International Application Number:
PCT/US2005/041639

(22) International Filing Date:
16 November 2005 (16.11.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/628,596 16 November 2004 (16.11.2004) US

(71) Applicant *(for all designated States except US)*: **AVIDIA RESEARCH INSTITUTE** [US/US]; 2450 Bayshore Boulevard, Mountain View, CA 94043 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **STEMMER, Willem, P., C.** [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US).

(74) Agents: **FANG, Carol, A.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

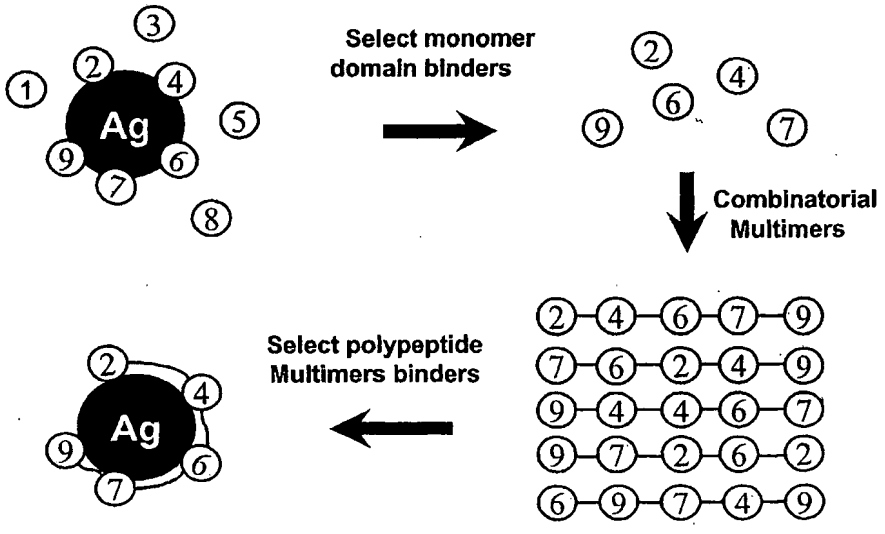
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN SCAFFOLDS AND USES THEROF



(57) Abstract: The present invention provides thrombospondin, thyroglobulin and trfoil/PD monomer domains and multimers comprising the monomer domains are provided. Methods, compositions, libraries and cells that express one or more library member, along with kits and integrated systems, are also included in the present invention.

WO 2006/127040 A2

PROTEIN SCAFFOLDS AND USES THEREOF

CROSS-REFERENCES TO RELATED APPLICATIONS

[01] The present application claims the benefit of U.S. Provisional Patent Application No. 60/628,596, filed November 16, 2004 and is a continuation in part of U.S.S.N. 10/871602, filed June 17, 2004, which is a continuation-in-part application of U.S.S.N. 10/840,723, filed May 5, 2004, which is a continuation-in-part application of U.S.S.N. 10/693,056, filed October 24, 2003 and a continuation-in-part of U.S.S.N. 10/693,057, filed October 24, 2003, both of which are continuations-in-part of U.S.S.N. 10/289,660, filed November 6, 2002, which is a continuation-in-part application of U.S.S.N. 10/133,128, filed April 26, 2002, which claims benefit of priority to U.S.S.N. 60/374,107, filed April 18, 2002, U.S.S.N. 60/333,359, filed November 26, 2001, U.S.S.N. 60/337,209, filed November 19, 2001, and U.S.S.N. 60/286,823, filed April 26, 2001, all of which are incorporated herein by reference in their entirety for all purposes.

BACKGROUND OF THE INVENTION

[02] Analysis of protein sequences and three-dimensional structures have revealed that many proteins are composed of a number of discrete monomer domains. Such proteins are often called 'mosaic proteins' because they are a linear mosaic of recurring building blocks. The majority of discrete monomer domain proteins is extracellular or constitutes the extracellular parts of membrane-bound proteins.

[03] An important characteristic of a discrete monomer domain is its ability to fold independently of the other domains in the same protein. Folding of these domains may require limited assistance from, *e.g.*, a chaperonin(s) (*e.g.*, a receptor-associated protein (RAP)), a metal ion(s), or a co-factor. The ability to fold independently prevents misfolding of the domain when it is inserted into a new protein or a new environment. This characteristic has allowed discrete monomer domains to be evolutionarily mobile. As a result, discrete domains have spread during evolution and now occur in otherwise unrelated proteins. Some domains, including the fibronectin type III domains and the immunoglobin-

like domain, occur in numerous proteins, while other domains are only found in a limited number of proteins.

[04] Proteins that contain these domains are involved in a variety of processes, such as cellular transporters, cholesterol movement, signal transduction and signaling functions which are involved in development and neurotransmission. *See* Herz, (2001) Trends in Neurosciences 24(4):193-195; Goldstein and Brown, (2001) Science 292: 1310-1312. The function of a discrete monomer domain is often specific but it also contributes to the overall activity of the protein or polypeptide. For example, the LDL-receptor class A domain (also referred to as a class A module, a complement type repeat or an A-domain) is involved in ligand binding while the gamma-carboxyglumatic acid (Gla) domain which is found in the vitamin-K-dependent blood coagulation proteins is involved in high-affinity binding to phospholipid membranes. Other discrete monomer domains include, *e.g.*, the epidermal growth factor (EGF)-like domain in tissue-type plasminogen activator which mediates binding to liver cells and thereby regulates the clearance of this fibrinolytic enzyme from the circulation and the cytoplasmic tail of the LDL-receptor which is involved in receptor-mediated endocytosis.

[05] Individual proteins can possess one or more discrete monomer domains. Proteins containing a large number of recurring domains are often called mosaic proteins. For example, members of the LDL-receptor family contain a large number of domains belonging to four major families: the cysteine rich A-domain repeats, epidermal growth factor precursor-like repeats, a transmembrane domain and a cytoplasmic domain. The LDL-receptor family includes members that: 1) are cell-surface receptors; 2) recognize extracellular ligands; and 3) internalize them for degradation by lysosomes. See Hussain *et al.*, (1999) Annu. Rev. Nutr. 19:141-72. For example, some members include very-low-density lipoprotein receptors (VLDL-R), apolipoprotein E receptor 2, LDLR-related protein (LRP) and megalin. Family members have the following characteristics: 1) cell-surface expression; 2) extracellular ligand binding mediated by A-domains; 3) requirement of calcium for folding and ligand binding; 4) recognition of receptor-associated protein and apolipoprotein (apo) E; 5) epidermal growth factor (EGF) precursor homology domain containing YWTD repeats; 6) single membrane-spanning region; and 7) receptor-mediated endocytosis of various ligands. *See* Hussain, *supra*. These family members bind several structurally dissimilar ligands.

[06] It is advantageous to develop methods for generating and optimizing the desired properties of these discrete monomer domains. However, the discrete monomer

domains, while often being structurally conserved, are not conserved at the nucleotide or amino acid level, except for certain amino acids, *e.g.*, the cysteine residues in the A-domain. Thus, existing nucleotide recombination methods fall short in generating and optimizing the desired properties of these discrete monomer domains.

[07] The present invention addresses these and other problems.

BRIEF SUMMARY OF THE INVENTION

[08] The present invention provide proteins comprising monomer domains that specifically bind to target molecules, polynucleotides encoding the proteins, methods of using such proteins, methods of identifying monomer domains for use in such proteins, and libraries comprising monomer domains.

[09] One embodiment of the invention provides proteins comprising a non-naturally occurring monomer domain that specifically binds to a target molecule. The monomer domain is 30-100 amino acids in length and is selected from a thrombospondin monomer domain and a thyroglobulin monomer domain. In some embodiments, the the monomer domain comprises at least one, two, three, or more disulfide bonds In some embodiments, C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds and C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds. In some embodiments, the thrombospondin monomer domain sequence comprises no more than three point insertions, mutations, or deletions from the following sequence:

(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; and the thyroglobulin monomer domain comprises no more than three point insertions, mutations, or deletions from the following sequence:

C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6; wherein "x" is any amino acid. In some embodiments, the thrombospondin monomer domain comprises the following sequence:

(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; and the thyroglobulin monomer domain comprises n the following sequence:

C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6; wherein "x" is any amino acid. In some embodiments, the thrombospondin monomer domain sequence comprises no more than three point insertions, mutations, or deletions from the following sequence:

(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]x[C_3vldr]xxxx([Pq]xx)xxxxx[C_4ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds; the thyroglobulin monomer domain sequence comprises no more than three point insertions, mutations, or deletions from the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC$_3$x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; α is selected from: w, y, f, and l; ϕ is selected from: d, e, and n; and "x" is selected from any amino acid. In some embodiments, the thrombospondin monomer domain comprises the following sequence:

(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]x[C_3vldr]xxxx([Pq]xx)xxxxx[C_4ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds; the thyroglobulin monomer domain comprises the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC$_3$x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and α is selected from: w, y, f, and l; ϕ is selected from: d, e, and n; and "x" is selected from any amino acid. In some embodiments, the thrombospondin monomer domain sequence comprises no more than three point insertions, mutations, or deletions from the following sequence:

C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6; the thyroglobulin monomer domain sequence comprises no more than three point insertions, mutations, or deletions from the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC$_3$x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6. In some embodiments, the thrombospondin monomer comprises the following sequence:

C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6; and the thyroglobulin monomer domain sequence comprises the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC$_3$x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6.

[10] The invention also provides a protein, comprising a non-naturally occurring monomer domain that specifically binds to a target molecule. The target molecule is not bound by a naturally-occurring monomer domain that is at least 75%, 80%, 85%, 90%, 85%, 98%, or 99% identical to the non-naturally occurring monomer domain and the non-

naturally occurring monomer domain is selected from a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain. In some embodiments, the monomer domain comprises at least one, two, three, or more disulfide bonds. In some embodiments, the monomer domain is 30-100 amino acids in length. In some embodiments, the thrombospondin monomer domain comprises the following sequence: (wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; the trefoil monomer domain comprises the following sequence: C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and the thyroglobulin monomer domain comprises the following sequence: C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6 and "x" is any amino acid. In some embodiments, C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds; and C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds. In some embodiments, the thrombospondin monomer domain comprises the following sequence: (WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]x[C_3vldr]xxxx([Pq]xx)xxxxx[C_4ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds; the trefoil monomer domain comprises the following sequence: C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C_3xx[krln][Gnk]C_4C_5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy]; the thyroglobulin monomer domain comprises the following sequence: C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC_3x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[φg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and α is selected from: w, y, f, and l; φ is selected from: d, e, and n; and "x" is selected from any amino acid. In some embodiments, the thrombospondin monomer comprises the following sequence: C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6; the trefoil monomer domain comprises the following sequence: C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C_6; and the thyroglobulin monomer comprises the following sequence:

C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6.

[11] The invention further provides a composition comprising at least two monomer domains, wherein at least one monomer domain is a non-naturally occurring monomer domain and the monomer domains bind an ion and at least one monomer domain is selected from: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain. In some embodiments, at least one of the two monomer domains is less than about 50 kD. In some embodiments, the two domains are linked by a peptide linker. In some embodiments, wherein the linker is heterologous to at least one of the monomer domains. In some embodiments, the thrombospondin monomer domain comprises the following sequence:
(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; he trefoil monomer domain comprises the following sequence:
C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and the thyroglobulin monomer domain comprises the following sequence:
C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vdxx(x)GxxxxGxxxxxgxx(xx)xC_6; and "x" is any amino acid. In some embodiments, C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds; and C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds. In some embodiments, the thrombospondin monomer domain comprises the following sequence:
(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]x[C_3vldr]xxxx([Pq]xx)xxxxx[C_4ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds; the trefoil monomer domain comprises the following sequence:
C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C_3xx[krln][Gnk]C_4C_5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy]; the thyroglobulin monomer domain comprises the following sequence:
C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC_3x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and and α is selected from: w, y, f, and l; ϕ is selected from: d, e, and n; and "x" is selected from any amino acid. In some embodiments, the thrombospondin monomer comprises the following sequence:
C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6; the trefoil monomer domain comprises the following

sequence:
C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C_6; and the thyroglobulin monomer comprises the following sequence:
C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6.

[12] The invention further provides isolated polynucleotides encoding the proteins described herein and cells comprising the polynucleotides.

[13] The invention also provides methods for identifying a monomer domain that binds to a target molecule by: (1) providing a library of non-naturally-occurring monomer domains, wherein the monomer domain is selected from: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain, wherein the thrombospondin monomer domain comprises the following sequence:
(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; the trefoil monomer domain comprises the following sequence:
C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and the thyroglobulin monomer domain comprises the following sequence:
C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vdxx(x)GxxxxGxxxxxgxx(xx)xC_6; and "x" is any amino acid; (2) screening the library of monomer domains for affinity to a first target molecule; and (3)identifying at least one monomer domain that binds to at least one target molecule. In some embodiments, the at least one monomer domain specifically binds to a target molecule that is not bound by a naturally-occurring monomer domain that is at least 90% identical to the non-naturally occurring monomer domain. In some embodiments, C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds; and C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds. In some embodiments, the thrombospondin monomer domain comprises the following sequence:
(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]x[C_3vldr]xxxx([Pq]xx)xxxxx[C_4ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds; the trefoil monomer domain comprises the following sequence:
C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C_3xx[krln][Gnk]C_4C_5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy]; the thyroglobulin monomer domain

comprises the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC_3x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and α is selected from: w, y, f, and l; ϕ is selected from: d, e, and n; and "x" is selected from any amino acid. In some embodiments, the thrombospondin monomer comprises the following sequence:

C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6; the trefoil monomer domain comprises the following sequence:

C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C_6; and the thyroglobulin monomer comprises the following sequence:

C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6. In some embodiments, the method further comprises linking the identified monomer domains to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains; screening the library of multimers for the ability to bind to the first target molecule; and identifying a multimer that binds to the first target molecule. Each monomer domain of the selected multimer binds to the same target molecule or to different target molecules. In some embodiments, the selected multimer comprises two, three, four, or more monomer domains. In some embodiments, the methods further comprises a step of mutating at least one monomer domain, thereby providing a library comprising mutated monomer domains. In some embodiments, the mutating step comprises recombining a plurality of polynucleotide fragments of at least one polynucleotide encoding a polypeptide domain. In some embodiments, the methods further comprises screening the library of monomer domains for affinity to a second target molecule; identifying a monomer domain that binds to a second target molecule; linking at least one monomer domain with affinity for the first target molecule with at least one monomer domain with affinity for the second target molecule, thereby forming a multimer with affinity for the first and the second target molecule. In some embodiments, the library of monomer domains is expressed as a phage display, ribosome display or cell surface display. In some embodiments, the library of monomer domains is presented on a microarray.

[14] The invention further comprises a library of proteins comprising non-naturally-occurring monomer domains, wherein the monomer domain is selected from: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain. In some embodiments, the thrombospondin monomer domain comprises the following sequence: (wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6; the trefoil monomer domain comprises the following sequence: C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and the thyroglobulin monomer domain comprises the following sequence: C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxP xC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6; and "x" is any amino acid. In some embodiments,each monomer domain of the multimers is a non-naturally occurring monomer domain. In some embodiments, the library comprises a plurality of multimers, wherein the multimers comprise at least two monomer domains linked by a linker. In some embodiments, the library comprises at least 100 different proteins comprising different monomer domains.

[15] The present invention also provides methods for identifying domain monomers and multimers that bind to a target molecule. In some embodiments, the method comprises: providing a library of monomer domains; screening the library of monomer domains for affinity to a first target molecule; and identifying at least one monomer domain that binds to at least one target molecule. In some embodiments, the monomer domains each bind an ion (*e.g.*, calcium).

[16] In some embodiments, the methods further comprise linking the identified monomer domains to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains; screening the library of multimers for the ability to bind to the first target molecule; and identifying a multimer that binds to the first target molecule.

[17] In some embodiments, each monomer domain of the selected multimer binds to the same target molecule. In some embodiments, the selected multimer comprises three monomer domains. In some embodiments, the selected multimer comprises four monomer domains.

[18] In some embodiments, the monomer domains are selected from a Thrombospondin type I domain, a thyroglobulin type I repeat domain, a Trefoil (P-type) domain, and an EGF-like domain (*e.g.*, a Laminin-type EGF-like domain).

[19] In some embodiments, the methods comprise a further step of mutating at least one monomer domain, thereby providing a library comprising mutated monomer domains. In some embodiments, the mutating step comprises recombining a plurality of polynucleotide fragments of at least one polynucleotide encoding a monomer domain. In some embodiments, the mutating step comprises directed evolution; combining different loop sequences; site-directed mutagenesis; or site-directed recombination to create crossovers that result in the generation of sequences that are identical to human sequences.

[20] In some embodiments, the methods further comprise: screening the library of monomer domains for affinity to a second target molecule; identifying a monomer domain that binds to a second target molecule; linking at least one monomer domain with affinity for the first target molecule with at least one monomer domain with affinity for the second target molecule, thereby forming a multimer with affinity for the first and second target molecule.

[21] In some embodiments, the target molecule is selected from a viral antigen, a bacterial antigen, a fungal antigen, an enzyme, a cell surface protein, an intracellular protein, an enzyme inhibitor, a reporter molecule, a serum protein, and a receptor. In some embodiments, the viral antigen is a polypeptide required for viral replication.

[22] In some embodiments, the library of monomer domains is expressed as by phage display, phagemid display, ribosome display, polysome display, or cell surface display (*e.g.*, *E. coli* cell surface display), yeast cell surface display or display via fusion to a protein that binds to the polynucleotide encoding the protein. In some embodiments, the library of monomer domains is presented on a microarray, including 96-well, 384 well or higher density microtiter plates.

[23] In some embodiments, the monomer domains are linked by a polypeptide linker. In some embodiments, the polypeptide linker is a linker naturally-associated with the monomer domain. In some embodiments, the polypeptide linker is a linker naturally-associated with the family of monomer domains. In some embodiments, the polypeptide linker is a variant of a linker naturally-associated with the monomer domain. In some embodiments the linker is a gly-ser linker. In some embodiments, the linking step comprises linking the monomer domains with a variety of linkers of different lengths and composition.

[24] In some embodiments, the domains form a secondary and tertiary structure by the formation of disulfide bonds. In some embodiments, the multimers comprise

an A domain connected to a monomer domain by a polypeptide linker. In some embodiments, the linker is from 1-20 amino acids inclusive. In some embodiments, the linker is made up of 5-7 amino acids. In some embodiments, the linker is 6 amino acids in length. In some embodiments, the linker comprises the following sequence, $A_1A_2A_3A_4A_5A_6$, wherein A_1 is selected from the amino acids A, P, T, Q, E and K; A_2 and A_3 are any amino acid except C, F, Y, W, or M; A_4 is selected from the amino acids S, G and R; A_5 is selected from the amino acids H, P, and R; A_6 is the amino acid, T. In some embodiments, the linker comprises a naturally-occurring sequence between the C-terminal cysteine of a first A domain and the N-terminal cysteine of a second A domain. In some embodiments the linker comprises glycine and serine.

[25] The present invention also provides methods for identifying a multimer that binds to at least one target molecule, comprising the steps of: providing a library of multimers, wherein each multimer comprises at least two monomer domains and wherein each monomer domain exhibits a binding specificity for a target molecule; and screening the library of multimers for target molecule-binding multimers. In some embodiments, the methods further comprise identifying target molecule-binding multimers having an avidity for the target molecule that is greater than the avidity of a single monomer domain for the target molecule. In some embodiments, one or more of the multimers comprises a monomer domain that specifically binds to a second target molecule.

[26] Alternative methods for identifying a multimer that binds to a target molecule include methods comprising providing a library of monomer domains and/or immuno domains; screening the library of monomer domains and/or immuno domain for affinity to a first target molecule; identifying at least one monomer domain and/or immuno domain that binds to at least one target molecule; linking the identified monomer domain and/or immuno domain to a library of monomer domains and/or immuno domains to form a library of multimers, each multimer comprising at least two monomer domains, immuno domains or combinations thereof; screening the library of multimers for the ability to bind to the first target molecule; and identifying a multimer that binds to the first target molecule.

[27] In some embodiments, the monomer domains each bind an ion. In some embodiments, the ion is selected from calcium and zinc.

[28] In some embodiments, the linker comprises at least 3 amino acid residues. In some embodiments, the linker comprises at least 6 amino acid residues. In some embodiments, the linker comprises at least 10 amino acid residues.

[29] The present invention also provides polypeptides comprising at least two monomer domains separated by a heterologous linker sequence. In some embodiments, each monomer domain specifically binds to a target molecule; and each monomer domain is a non-naturally occurring protein monomer domain. In some embodiments, each monomer domain binds an ion.

[30] In some embodiments, polypeptides comprise a first monomer domain that binds a first target molecule and a second monomer domain that binds a second target molecule. In some embodiments, the polypeptides comprise two monomer domains, each monomer domain having a binding specificity that is specific for a different site on the same target molecule. In some embodiments, the polypeptides further comprise a monomer domain having a binding specificity for a second target molecule.

[31] In some embodiments, the monomer domains of a library, multimer or polypeptide are typically about 40% identical to each other, usually about 50% identical, sometimes about 60% identical, and frequently at least 70% identical.

[32] The invention also provides polynucleotides encoding the above-described polypeptides.

[33] The present invention also provides multimers of immuno-domains having binding specificity for a target molecule, as well as methods for generating and screening libraries of such multimers for binding to a desired target molecule. More specifically, the present invention provides a method for identifying a multimer that binds to a target molecule, the method comprising, providing a library of immuno-domains; screening the library of immuno-domains for affinity to a first target molecule; identifying one or more (*e.g.*, two or more) immuno-domains that bind to at least one target molecule; linking the identified monomer domain to form a library of multimers, each multimer comprising at least three immuno-domains (*e.g.*, four or more, five or more, six or more, *etc.*); screening the library of multimers for the ability to bind to the first target molecule; and identifying a multimer that binds to the first target molecule. Libraries of multimers of at least two immuno-domains that are minibodies, single domain antibodies, Fabs, or combinations thereof are also employed in the practice of the present invention. Such libraries can be readily screened for multimers that bind to desired target molecules in accordance with the invention methods described herein.

[34] The present invention further provides methods of identifying hetero-immuno multimers that binds to a target molecule. In some embodiments, the methods comprise, providing a library of immuno-domains; screening the library of immuno-domains

for affinity to a first target molecule; providing a library of monomer domains; screening the library of monomer domains for affinity to a first target molecule; identifying at least one immuno-domain that binds to at least one target molecule; identifying at least one monomer domain that binds to at least one target molecule; linking the identified immuno-domain with the identified monomer domains to form a library of multimers, each multimer comprising at least two domains; screening the library of multimers for the ability to bind to the first target molecule; and identifying a multimer that binds to the first target molecule.

[35] The present invention also provides methods for identifying a laminin-EGF monomer domain, a thrombospondin type I monomer domain, a thyroglobulin monomer domain, or a trefoil monomer domain that binds to a target molecule. In some embodiments, the method comprises providing a library of laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains; screening the library of laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains for affinity to a target molecule; and identifying a laminin-EGF monomer domain, thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain that binds to the target molecule.

[36] In some embodiments, the method comprises linking each member of a library of laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains to the identified monomer domain to form a library of multimers; screening the library of multimers for affinity to the target molecule; and identifying a multimer that binds to the target. In some embodiments, the multimer binds to the target with greater affinity than the monomer. In some embodiments, the method further comprises expressing the library using a display format selected from a phage display, a ribosome display, a polysome display, or a cell surface display.

[37] In some embodiments, the method further comprises a step of mutating at least one monomer domain, thereby providing a library comprising mutated laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains. In some embodiments, the mutating step comprises directed evolution; site-directed mutagenesis; by combining different loop sequences, or by site-directed recombination to create crossovers that result in generation of sequences that are identical to human sequences.

[38] The present invention also provides method of producing a polypeptide comprising the multimer identified in a method comprising providing a library of laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains; screening the library of laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains for affinity to a target molecule; and identifying a laminin-EGF monomer domain, thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain that binds to the target molecule. In some embodiments, the multimer is produced by recombinant gene expression.

[39] The present invention also provides methods for generating a library of thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains derived from thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains. In some embodiments, the methods comprise providing loop sequences corresponding to at least one loop from each of two different naturally occurring variants of a human laminin-EGF monomer domains, thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains, wherein the loop sequences are polynucleotide or polypeptide sequences; covalently combining loop sequences to generate a library of chimeric monomer domain sequences, each chimeric sequence encoding a chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain having at least two loops; expressing the library of chimeric thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains using a display format selected from phage display, ribosome display, polysome display, and cell surface display; screening the expressed library of chimeric thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains for binding to a target molecule; and identifying a chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain that binds to the target molecule.

[40] In some embodiments, the methods further comprise linking the identified chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain to each member of the library of chimeric thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains to form a library of multimers; screening the library of multimers for the ability to bind to the first target molecule with an increased affinity; and identifying a multimer of chimeric thrombospondin type I monomer domains, thyroglobulin monomer

domains, or trefoil monomer domains that binds to the first target molecule with an increased affinity.

[41] The present invention also provides methods of making chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain identified in a method comprising providing loop sequences corresponding to at least one loop from each of two different naturally occurring variants of a human thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains, wherein the loop sequences are polynucleotide or polypeptide sequences; covalently combining loop sequences to generate a library of chimeric monomer domain sequences, each chimeric sequence encoding a chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain having at least two loops; expressing the library of chimeric thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains using a display format selected from phage display, ribosome display, polysome display, and cell surface display; screening the expressed library of chimeric thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains for binding to a target molecule; and identifying a chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain that binds to the target molecule. In some embodiments, the chimeric thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain is produced by recombinant gene expression.

[42] In some embodiments, the monomer domain binds to a target molecule. In some embodiments, the polypeptide is 45 or fewer amino acids long. In some embodiments, the heterologous amino acid sequence is selected from an affinity peptide, a heterologous thrombospondin type I monomer domain, a heterologous thyroglobulin monomer domain, or a heterologous trefoil monomer domain, a purification tag, an enzyme (*e.g.*, horseradish peroxidase or alkaline phosphatase), and a reporter protein (*e.g.*, green fluorescent protein or luciferase). In some embodiments, the target is not a variable region or hypervariable region of an antibody.

[43] The present invention provides methods for screening a library of monomer domains or multimers comprising monomer domains for binding affinity to multiple ligands. In some embodiments, the method comprises contacting a library of monomer domains or multimers of monomer domains to multiple ligands; and selecting monomer domains or multimers that bind to at least one of the ligands.

[44] In some embodiments, the methods comprise (i.) contacting a library of monomer domains to multiple ligands; (ii.) selecting monomer domains that bind to at least one of the ligands; (iii.) linking the selected monomer domains to a library of monomer domains to form a library of multimers, each comprising a selected monomer domain and a second monomer domain; (iv.) contacting the library of multimers to the multiple ligands to form a plurality of complexes, each complex comprising a multimer and a ligand; and (v.) selecting at least one complex.

[45] In some embodiments, the method further comprises linking the multimers of the selected complexes to a library of monomer domains or multimers to form a second library of multimers, each comprising a selected multimer and at least a third monomer domain; contacting the second library of multimers to the multiple ligands to form a plurality of second complexes; and selecting at least one second complex.

[46] In some embodiments, the identity of the ligand and the multimer is determined. In some embodiments, a library of monomer domains is contacted to multiple ligands. In some embodiments, a library of multimers is contacted to multiple ligands.

[47] In some embodiments, the multiple ligands are in a mixture. In some embodiments, the multiple ligands are in an array. In some embodiments, the multiple ligands are in or on a cell or tissue. In some embodiments, the multiple ligands are immobilized on a solid support.

[48] In some embodiments, the ligands are polypeptides. In some embodiments, the polypeptides are expressed on the surface of phage. In some embodiments, the monomer domain or multimer library is expressed on the surface of phage.

[49] In some embodiments, the library of multimers is expressed on the surface of phage to form library-expressing phage and the ligands are expressed on the surface of phage to form ligand-expressing phage, and the method comprises contacting library-expressing phage to the ligand-expressing phage to form ligand-expressing phage/library-expressing phage pairs; removing ligand-expressing phage that do not bind to library-expressing or removing library-expressing phage that do not bind to ligand-expressing phage; and selecting the ligand-expressing phage/library-expressing phage pairs. In some embodiments, the methods further comprise isolating polynucleotides from the phage pairs and amplifying the polynucleotides to produce a polynucleotide hybrid comprising polynucleotides from the ligand-expressing phage and the library-expressing phage.

[50] In some embodiments, the methods comprise isolating polynucleotide hybrids from a plurality of phage pairs, thereby forming a mixture of polynucleotide hybrids.

In some embodiments, the methods comprise contacting the mixture of hybrid polynucleotides to a cDNA library under conditions to allow for polynucleotide hybridization, thereby hybridizing a hybrid polynucleotide to a cDNA in the cDNA library; and determining the nucleotide sequence of the hybridized hybrid polynucleotide, thereby identifying a monomer domain that specifically binds to the polypeptide encoded by the cDNA. In some embodiments, the monomer domain library is expressed on the surface of phage to form library-expressing phage and the ligands are expressed on the surface of phage to form ligand-expressing phage, and the selected complexes comprise a library-expressing phage bound to a ligand-expressing phage and the method comprises: dividing the selected monomer domains or multimers into a first and a second portion, linking the monomer domains or multimers of the first portion to a solid surface and contacting a phage-displayed ligand library to the monomer domains or multimers of the first portion to identify target ligand phage that binds to a monomer domain or multimer of the first portion; infecting phage displaying the monomer domains or multimers of the second portion into bacteria to express the phage; and contacting the target ligand phage to the expressed phage to form phage pairs comprised of a target ligand phage and a phage displaying a monomer domain or multimer.

[51] In some embodiments, the methods further comprise isolating a polynucleotide from each phage of the phage pair, thereby identifying a multimer or monomer domain that binds to the ligand in the phage pair. In some embodiments, the methods further comprise amplifying the polynucleotides to produce a polynucleotide hybrid comprising polynucleotides from the target ligand phage and the library phage.

[52] In some embodiments, the methods comprise isolating and amplifying polynucleotide hybrids from a plurality of phage pairs, thereby forming a mixture of polynucleotide hybrids. In some embodiments, the methods comprise contacting the mixture of hybrid polynucleotides to a cDNA library under conditions to allow for hybridization, thereby hybridizing a hybrid polynucleotide to a cDNA in the cDNA library; and determining the nucleotide sequence of the associated hybrid polynucleotide, thereby identifying a monomer domain that specifically binds to the ligand encoded by the cDNA associated cDNA.

[53] The present invention also provides non-naturally-occurring polypeptides comprising an amino acid sequence in which:

at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20% or more of the amino acids in the sequence are cysteine; and

the amino acid sequence is at least 10, 20, 30, 45, 50, 55, 60, 70, 80, 90, 100 or more amino acids long; and/or

the amino acid sequence is less than 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, or 40 amino acids long;and/or

at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% or more of the amino acids are non-naturally-occurring amino acids. For example, in some embodiments, the amino acid sequence comprises at least 10% cysteines and the amino acid sequence is at least 50 amino acids long or at least 25% of the amino acids are non-naturally occurring. In some embodiments, the amino acid sequence is a non-naturally occurring A domain.

[54] In some embodiments, the polypeptides of the invention comprise one, two, three, four, or more monomers with at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% or more non-naturally-occurring amino acids. In some embodiments, the one or more monomer domains comprises at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% or more amino acids that do not occur at that position in natural human proteins. In some embodiments, the monomer domains are derived from a naturally-occurring human protein sequence. In some embodiments, the polypeptides of the invention also have a serum half-life of at least, *e.g.*, 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70 80, 90, 100, 150, 200, 250, 400, 500 or more hours.

DEFINITIONS

[55] Unless otherwise indicated, the following definitions supplant those in the art.

[56] The term "monomer domain" or "monomer" is used interchangeably herein refer to a discrete region found in a protein or polypeptide. A monomer domain forms a native three-dimensional structure in solution in the absence of flanking native amino acid sequences. Monomer domains of the invention can be selected to specifically bind to a target molecule. As used herein, the term "monomer domain" does not encompass the complementarity determining region (CDR) of an antibody.

[57] The term "monomer domain variant" refers to a domain resulting from human-manipulation of a monomer domain sequence. Examples of man-manipulated changes include, *e.g.*, random mutagenesis, site-specific mutagenesis, recombining, directed evolution, oligo-directed forced crossover events, direct gene synthesis incorporation of mutation, *etc.* The term "monomer domain variant" does not embrace a mutagenized complementarity determining region (CDR) of an antibody.

[58] The term "loop" refers to that portion of a monomer domain that is typically exposed to the environment by the assembly of the scaffold structure of the monomer domain protein, and which is involved in target binding. The present invention provides three types of loops that are identified by specific features, such as, potential for disulfide bonding, bridging between secondary protein structures, and molecular dynamics (*i.e.*, flexibility). The three types of loop sequences are a cysteine-defined loop sequence, a structure-defined loop sequence, and a B-factor-defined loop sequence.

[59] As used herein, the term "cysteine-defined loop sequence" refers to a subsequence of a naturally occurring monomer domain-encoding sequence that is bound at each end by a cysteine residue that is conserved with respect to at least one other naturally occurring monomer domain of the same family. Cysteine-defined loop sequences are identified by multiple sequence alignment of the naturally occurring monomer domains, followed by sequence analysis to identify conserved cysteine residues. The sequence between each consecutive pair of conserved cysteine residues is a cysteine-defined loop sequence. The cysteine-defined loop sequence does not include the cysteine residues adjacent to each terminus. Monomer domains having cysteine-defined loop sequences include the thrombospondin domains, thyroglobulin domains, trefoil/PD domains, and the like. Thus, for example, thrombospondin domains are represented by the consensus sequence, $CX_3CX_{10}CX_{16}CX_{11}CX_4C$, wherein X_3, X_{10}, X_{16}, X_{11}, and X_4, each represent a cysteine-defined loop sequence; trefoil/PD domains are represented by the consensus sequence, $CX_{10}CX_9CX_4CCX_{10}C$, wherein X_{10}, X_9, X_4, and X_{10}, each represent a cysteine-defined loop sequence; and thyroglobulin domains are represented by the consensus sequence, $CX_{26}CX_{10}CX_6CX_1CX_{18}C$, wherein X_{26}, X_{10}, X_6, X_1, and X_{18}, each represent a cysteine-defined loop sequence.

[60] The term "multimer" is used herein to indicate a polypeptide comprising at least two monomer domains and/or immuno-domains (*e.g.*, at least two monomer domains, at least two immuno-domains, or at least one monomer domain and at least one immuno-domain). The separate monomer domains and/or immuno-domains in a

multimer can be joined together by a linker. A multimer is also known as a combinatorial mosaic protein or a recombinant mosaic protein.

[61] The term "family" and "family class" are used interchangeably to indicate proteins that are grouped together based on similarities in their amino acid sequences. These similar sequences are generally conserved because they are important for the function of the protein and/or the maintenance of the three dimensional structure of the protein. Examples of such families include the LDL Receptor A-domain family, the EGF-like family, and the like.

[62] The term "ligand," also referred to herein as a "target molecule," encompasses a wide variety of substances and molecules, which range from simple molecules to complex targets. Target molecules can be proteins, nucleic acids, lipids, carbohydrates or any other molecule capable of recognition by a polypeptide domain. For example, a target molecule can include a chemical compound (*i.e.*, non-biological compound such as, *e.g.*, an organic molecule, an inorganic molecule, or a molecule having both organic and inorganic atoms, but excluding polynucleotides and proteins), a mixture of chemical compounds, an array of spatially localized compounds, a biological macromolecule, a bacteriophage peptide display library, a polysome peptide display library, an extract made from a biological materials such as bacteria, plants, fungi, or animal (*e.g.*, mammalian) cells or tissue, a protein, a toxin, a peptide hormone, a cell, a virus, or the like. Other target molecules include, *e.g.*, a whole cell, a whole tissue, a mixture of related or unrelated proteins, a mixture of viruses or bacterial strains or the like. Target molecules can also be defined by inclusion in screening assays described herein or by enhancing or inhibiting a specific protein interaction (*i.e.*, an agent that selectively inhibits a binding interaction between two predetermined polypeptides).

[63] As used herein, the term "immuno-domains" refers to protein binding domains that contain at least one complementarity determining region (CDR) of an antibody. Immuno-domains can be naturally occurring immunological domains (*i.e.* isolated from nature) or can be non-naturally occurring immunological domains that have been altered by human-manipulation (*e.g.*, via mutagenesis methods, such as, for example, random mutagenesis, site-specific mutagenesis, recombination, and the like, as well as by directed evolution methods, such as, for example, recursive error-prone PCR, recursive recombination, and the like.). Different types of immuno-domains that are suitable for use in the practice of the present invention include a minibody, a single-domain antibody, a single chain variable fragment (ScFv), and a Fab fragment.

[64] The term "minibody" refers herein to a polypeptide that encodes only 2 complementarity determining regions (CDRs) of a naturally or non-naturally (*e.g.*, mutagenized) occurring heavy chain variable domain or light chain variable domain, or combination thereof. An example of a minibody is described by Pessi *et al.*, *A designed metal-binding protein with a novel fold,* (1993) Nature 362:367-369.

[65] As used herein, the term "single-domain antibody" refers to the heavy chain variable domain ("V_H") of an antibody, *i.e.*, a heavy chain variable domain without a light chain variable domain. Exemplary single-domain antibodies employed in the practice of the present invention include, for example, the Camelid heavy chain variable domain (about 118 to 136 amino acid residues) as described in Hamers-Casterman, C. *et al.*, *Naturally occurring antibodies devoid of light chains* (1993) Nature 363:446-448, and Dumoulin, *et al.*, *Single-domain antibody fragments with high conformational stability* (2002) Protein Science 11:500-515.

[66] The terms "single chain variable fragment" or "ScFv" are used interchangeably herein to refer to antibody heavy and light chain variable domains that are joined by a peptide linker having at least 12 amino acid residues. Single chain variable fragments contemplated for use in the practice of the present invention include those described in Bird, *et al.*, (1988) Science 242(4877):423-426 and Huston *et al.*, (1988) PNAS USA 85(16):5879-83.

[67] As used herein, the term "Fab fragment" refers to an immuno-domain that has two protein chains, one of which is a light chain consisting of two light chain domains (V_L variable domain and C_L constant domain) and a heavy chain consisting of two heavy domains (*i.e.*, a V_H variable and a C_H constant domain). Fab fragments employed in the practice of the present invention include those that have an interchain disulfide bond at the C-terminus of each heavy and light component, as well as those that do not have such a C-terminal disulfide bond. Each fragment is about 47 kD. Fab fragments are described by Pluckthun and Skerra, (1989) Methods Enzymol 178:497-515.

[68] The term "linker" is used herein to indicate a moiety or group of moieties that joins or connects two or more discrete separate monomer domains. The linker allows the discrete separate monomer domains to remain separate when joined together in a multimer. The linker moiety is typically a substantially linear moiety. Suitable linkers include polypeptides, polynucleic acids, peptide nucleic acids and the like. Suitable linkers also include optionally substituted alkylene moieties that have one or more oxygen atoms incorporated in the carbon backbone. Typically, the molecular weight of the linker is less

than about 2000 daltons. More typically, the molecular weight of the linker is less than about 1500 daltons and usually is less than about 1000 daltons. The linker can be small enough to allow the discrete separate monomer domains to cooperate, *e.g.*, where each of the discrete separate monomer domains in a multimer binds to the same target molecule via separate binding sites. Exemplary linkers include a polynucleotide encoding a polypeptide, or a polypeptide of amino acids or other non-naturally occurring moieties. The linker can be a portion of a native sequence, a variant thereof, or a synthetic sequence. Linkers can comprise, *e.g.*, naturally occurring, non-naturally occurring amino acids, or a combination of both.

[69] The term "separate" is used herein to indicate a property of a moiety that is independent and remains independent even when complexed with other moieties, including for example, other monomer domains. A monomer domain is a separate domain in a protein because it has an independent property that can be recognized and separated from the protein. For instance, the ligand binding ability of the A-domain in the LDLR is an independent property. Other examples of separate include the separate monomer domains in a multimer that remain separate independent domains even when complexed or joined together in the multimer by a linker. Another example of a separate property is the separate binding sites in a multimer for a ligand.

[70] As used herein, "directed evolution" refers to a process by which polynucleotide variants are generated, expressed, and screened for an activity (*e.g.*, a polypeptide with binding activity) in a recursive process. One or more candidates in the screen are selected and the process is then repeated using polynucleotides that encode the selected candidates to generate new variants. Directed evolution involves at least two rounds of variation generation and can include 3, 4, 5, 10, 20 or more rounds of variation generation and selection. Variation can be generated by any method known to those of skill in the art, including, *e.g.*, by error-prone PCR, gene recombination, chemical mutagenesis and the like.

[71] The term "shuffling" is used herein to indicate recombination between non-identical sequences. In some embodiments, shuffling can include crossover via homologous recombination or via non-homologous recombination, such as via cre/lox and/or flp/frt systems. Shuffling can be carried out by employing a variety of different formats, including for example, *in vitro* and *in vivo* shuffling formats, in silico shuffling formats, shuffling formats that utilize either double-stranded or single-stranded templates, primer based shuffling formats, nucleic acid fragmentation-based shuffling formats, and oligonucleotide-mediated shuffling formats, all of which are based on recombination events

between non-identical sequences and are described in more detail or referenced herein below, as well as other similar recombination-based formats. The term "random" as used herein refers to a polynucleotide sequence or an amino acid sequence composed of two or more amino acids and constructed by a stochastic or random process. The random polynucleotide sequence or amino acid sequence can include framework or scaffolding motifs, which can comprise invariant sequences.

[72] The term "pseudorandom" as used herein refers to a set of sequences, polynucleotide or polypeptide, that have limited variability, so that the degree of residue variability at some positions is limited, but any pseudorandom position is allowed at least some degree of residue variation.

[73] The terms "polypeptide," "peptide," and "protein" are used herein interchangeably to refer to an amino acid sequence of two or more amino acids.

[74] "Conservative amino acid substitution" refers to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[75] The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids and DNA or RNA that performs a primarily structural role.

[76] The term "encoding" refers to a polynucleotide sequence encoding one or more amino acids. The term does not require a start or stop codon. An amino acid sequence can be encoded in any one of six different reading frames provided by a polynucleotide sequence.

[77] The term "promoter" refers to regions or sequence located upstream and/or downstream from the start of transcription that are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription.

[78] A "vector" refers to a polynucleotide, which when independent of the host chromosome, is capable of replication in a host organism. Examples of vectors include plasmids. Vectors typically have an origin of replication. Vectors can comprise, *e.g.*, transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular nucleic acid.

[79] The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[80] The phrase "specifically (or selectively) binds" to a polypeptide, when referring to a monomer or multimer, refers to a binding reaction that can be determinative of the presence of the polypeptide in a heterogeneous population of proteins and other biologics. Thus, under standard conditions or assays used in antibody binding assays, the specified monomer or multimer binds to a particular target molecule above background (*e.g.*, 2X, 5X, 10X or more above background) and does not bind in a significant amount to other molecules present in the sample.

[81] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same. "Substantially identical" refers to two or more nucleic acids or polypeptide sequences having a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region, or, when not specified, over the entire sequence), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Optionally, the identity or substantial identity exists over a region that is at least about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides or amino acids in length.

[82] A polynucleotide or amino acid sequence is "heterologous to" a second sequence if the two sequences are not linked in the same manner as found in naturally-occurring sequences. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence which is different from any naturally-occurring allelic

variants. The term "heterologous linker," when used in reference to a multimer, indicates that the multimer comprises a linker and a monomer that are not found in the same relationship to each other in nature (*e.g.*, they form a fusion protein).

[83] A "non-naturally-occurring amino acid" in a protein sequence refers to any amino acid other than the amino acid that occurs in the corresponding position in an alignment with a naturally-occurring polypeptide with the lowest smallest sum probability where the comparison window is the length of the monomer domain queried and when compared to the non-redundant ("nr") database of Genbank using BLAST 2.0 as described herein.

[84] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[85] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

[86] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The

sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[87] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g.,* Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995 supplement)).

[88] One example of a useful algorithm is the BLAST 2.0 algorithm, which is described in Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN

program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[89] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

BRIEF DESCRIPTION OF THE DRAWINGS

[90] Figure 1 schematically illustrates a general scheme for identifying monomer domains that bind to a ligand, isolating the selected monomer domains, creating multimers of the selected monomer domains by joining the selected monomer domains in various combinations and screening the multimers to identify multimers comprising more than one monomer that binds to a ligand.

[91] Figure 2 is a schematic representation of another selection strategy (guided selection). A monomer domain with appropriate binding properties is identified from a library of monomer domains. The identified monomer domain is then linked to monomer domains from another library of monomer domains to form a library of multimers. The multimer library is screened to identify a pair of monomer domains that bind simultaneously to the target. This process can then be repeated until the optimal binding properties are obtained in the multimer.

[92] Figure 3 illustrates walking selection to generate multimers that bind a target or targets with increased affinity.

[93] Figure 4 illustrates screening a library of monomer domains against multiple ligands displayed on a cell.

[94] Figure 5 illustrates monomer domain and multimer embodiments for increased avidity. While the figure illustrates specific gene products and binding affinities, it is appreciated that these are merely examples and that other binding targets can be used with the same or similar conformations.

[95] Figure 6 illustrates monomer domain and multimer embodiments for increased avidity. While the figure illustrates specific gene products and binding affinities, it is appreciated that these are merely examples and that other binding targets can be used with the same or similar conformations.

[96] Figure 7 illustrates various possible antibody- monomer or multimer of the invention) conformations. In some embodiments, the monomer or multimer replaces the Fab fragment of the antibody.

[97] Figure 8 illustrates a method for intradomain optimization of monomers.

[98] Figure 9 illustrates a possible sequence of multimer optimization steps in which optimal monomers and then multimers are selected followed by optimization of monomers, optimization of linkers and then optimization of multimers.

[99] Figure 10 illustrates four exemplary methods to recombine monomer and/or multimer libraries to introduce new variation. Figure 10A illustrates one exemplary embodiment of intra-domain recombination of monomers whereby portions of different monomers are recombined to form new monomers. Figure 10B illustrates a second embodiment of intra-domain recombination whereby portions of monomers recombined as set forth in Figure 10A are further recombined to form additional new monomers. Figure 10C illustrates one embodiment of inter-domain recombination, whereby different recombined monomers are linked to each other, *i.e.*, to form multimers. Figure 10D illustrates one embodiment of inter-module recombination whereby linked recombined monomers, *i.e.*, multimers that bind to the same target molecule are linked to other recombined monomers that recognize a different target molecule to form new multimers that simultaneously bind to different target molecules.

[100] Figure 11 depicts a possible conformation of a multimer of the invention comprising at least one monomer domain that binds to a half-life extending molecule and other monomer domains binding to two other different molecules. In the Figure, two monomer domains bind to a first target molecule and a separate monomer domain binds to a second target molecule.

DETAILED DESCRIPTION OF THE INVENTION

[101] The invention provides affinity agents comprising monomer domains, as well as multimers of the monomer domains. The affinity agents can be selected for the ability to bind to a desired ligand or mixture of ligands. The monomer domains and multimers can be screened to identify those that have an improved characteristic such as improved avidity or affinity or altered specificity for the ligand or the mixture of ligands, compared to the discrete monomer domain. The monomer domains of the present invention include specific variants of the laminin EGF-like domains, the thrombospondin Type 1 domains, the trefoil domains, and the thyroglobulin domains.

I. Monomer Domains

[102] Many suitable monomer domains can be used in the polypeptides of the invention. Typically suitable monomer domains comprise three disulfide bonds, 30 to 100 amino acids and have a binding site for a divalent metal ion, such as, *e.g.*, calcium. In some embodiments, thrombospondin type 1 monomer domains, trefoil monomer domains, or thyroglobulin monomer domains are used in the scaffolds of the invention. In other embodiments, laminin-EGF monomer domains are used.

[103] Monomer domains can have any number of characteristics. For example, in some embodiments, the monomer domains have low or no immunogenicity in an animal (*e.g.*, a human). Monomer domains can have a small size. In some embodiments, the monomer domains are small enough to penetrate skin or other tissues. Monomer domains can have a range of *in vivo* half-lives or stabilities. Characteristics of a monomer domain include the ability to fold independently and the ability to form a stable structure.

[104] Monomer domains can be polypeptide chains of any size. In some embodiments, monomer domains have about 25 to about 500, about 30 to about 200, about 30 to about 100, about 35 to about 50, about 35 to about 100, about 90 to about 200, about 30 to about 250, about 30 to about 60, about 9 to about 150, about 100 to about 150, about 25 to about 50, or about 30 to about 150 amino acids. Similarly, a monomer domain of the present invention can comprise, *e.g.*, from about 30 to about 200 amino acids; from about 25 to about 180 amino acids; from about 40 to about 150 amino acids; from about 50 to about 130 amino acids; or from about 75 to about 125 amino acids. Monomer domains and immuno-domains can typically maintain a stable conformation in solution, and are often heat stable, *e.g.*, stable at 95° C for at least 10 minutes without losing binding affinity. Monomer domains typically

bind with a K_d of less than about 10^{-15}, 10^{-14}, 10^{-13}, 10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.01 μM, about 0.1 μM , or about 1 μM. Sometimes, monomer domains and immuno-domains can fold independently into a stable conformation. In one embodiment, the stable conformation is stabilized by metal ions. The stable conformation can optionally contain disulfide bonds (*e.g.*, at least one, two, or three or more disulfide bonds). The disulfide bonds can optionally be formed between two cysteine residues. In some embodiments, monomer domains, or monomer domain variants, are substantially identical to the sequences exemplified (*e.g.*, thrombospondin, trefoil, or thyroglobulin) or otherwise referenced herein.

[105] Exemplary monomer domains that are particularly suitable for use in the practice of the present invention are cysteine-rich domains comprising disulfide bonds. Typically, the disulfide bonds promote folding of the domain into a three-dimensional structure. Usually, cysteine-rich domains have at least two disulfide bonds, more typically at least three disulfide bonds. Suitable cysteine rich monomer domains include, *e.g.*, the thrombospondin type 1 domain, the trefoil domain, or the thyroglobulin domain.

[106] The monomer domains can also have a cluster of negatively charged residues. Monomer domains may bind ion to maintain their secondary structure. Such monomer domains include, *e.g.*, A domains, EGF domains, EF Hand (*e.g.*, those present in calmodulin and troponin C), Cadherin domains, C-type lectins, C2 domains, Annexin, Gla-domains, Thrombospondin type 3 domains, all of which bind calcium, and zinc fingers (*e.g.*, C2H2 type C3HC4 type (RING finger), Integrase Zinc binding domain, PHD finger, GATA zinc finger, FYVE zinc finger, B-box zinc finger), which bind zinc. Without intending to limit the invention, it is believed that ion-binding stabilizes secondary structure while providing sufficient flexibility to allow for numerous binding conformations depending on primary sequence.

[107] The structure of the monomer domain is often conserved, although the polynucleotide sequence encoding the monomer need not be conserved. For example, domain structure may be conserved among the members of the domain family, while the domain nucleic acid sequence is not. Thus, for example, a monomer domain is classified as an Thrombospondin type 1 domain, a trefoil domain, or a thyroglobulin domain by its cysteine residues and its affinity for a metal ion (*e.g.*, calcium,) not necessarily by its nucleic acid sequence.

[108] In some embodiments, suitable monomer domains (*e.g.* domains with the ability to fold independently or with some limited assistance) can be selected from the

families of protein domains that contain β-sandwich or β-barrel three dimensional structures as defined by such computational sequence analysis tools as Simple Modular Architecture Research Tool (SMART), *see* Shultz *et al.*, *SMART: a web-based tool for the study of genetically mobile domains*, (2000) Nucleic Acids Research 28(1):231-234) or CATH (*see* Pearl *et.al.*, *Assigning genomic sequences to CATH*, (2000) Nucleic Acids Research 28(1):277-282).

[109] In some embodiments, the monomer domains are modified to bind to substrates to enhance protein function, including, for example, enzymatic activity and/or substrate conversion.

[110] As described herein, monomer domains may be selected for the ability to bind to targets other than the target that a homologous naturally occurring domain may bind. Thus, in some embodiments, the invention provides monomer domains (and multimers comprising such monomers) that do not bind to the target or the class or family of target proteins that a homologous naturally occurring domain may bind.

[111] Each of the domains described herein employ exemplary motifs (*i.e.*, scaffolds). Certain positions are marked x, indicating that any amino acid can occupy the position. These positions can include a number of different amino acid possibilities, thereby allowing for sequence diversity and thus affinity for different target molecules. Use of brackets in motifs indicates alternate possible amino acids within a position (*e.g.*, "[ekq]" indicates that either E, K or Q may be at that position). Use of parentheses in a motif indicates that that the positions within the parentheses may be present or absent (*e.g.*, "([ekq])" indicates that the position is absent or either E, K, or Q may be at that position). When more than one "x" is used in parentheses (*e.g.*, "(xx)"), each x represents a possible position. Thus "(xx)" indicates that zero, one or two amino acids may be at that position(s), where each amino acid is independently selected from any amino acid. α represents an aromatic/hydrophobic amino acid such as, *e.g.*, W, Y, F, or L; β represents a hydrophobic amino acid such as, *e.g.*, V, I, L, A, M, or F; χ represents a smallor polar amino acid such as, *e.g.*, G, A, S, or T; δ represents a charged amino acid such as, *e.g.*, K, R, E, Q, or D; ε represents a small amino acid such as, *e.g.*,; V, A, S, or T; and ϕ represents a negatively charged amino acid such as, *e.g.*, D, E, or N.

[112] Suitable domains include, *e.g.* thrombospondin type 1 domains, trefoil domains, and thyroglobulin domains.

[113] Thrombospondin type 1 ("TSP1") domains contain about 30-50 or 30-65 amino acids. In some embodiments, the domains comprise about 35-55 amino acids and in some cases about 50 amino acids. Within the 35-55 amino acids, there are typically about 4 to about 6 cysteine residues. Of the six cysteines, disulfide bonds typically are found between the following cysteines: C1 and C5, C2 and C6, C3 and C4. The cysteine residues of the domain are disulfide linked to form a compact, stable, functionally independent moiety comprising distorted beta strands. Clusters of these repeats make up a ligand binding domain, and differential clustering can impart specificity with respect to the ligand binding.

[114] Exemplary TSP1 domain sequences and consensus sequences are as follows:

(1) (xxxxxx)C_1xxxC_2xxxxx(x)xxxxxC_3xxxx(xxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6;
(2) (wxxWxx)C_1xxxC_2xxGxx(x)xRxxxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6
(3) (wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6
(4)
(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]xC_3[vldr]xxxx([Pq]xx)xxxxx C_4[ldae]xxxxxx(x)xxxC_5(x)xxxxC_6;
(5)
(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]xC_3[vldr]xxxx([Pq]xx)xxxxx C_4[ldae]xxxxxx(x)xxxC_5(x)xxxxC_6; and
(6)
C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxx xx)C_4xxxxxxxxx(xx)C_5xxxxC_6

[115] In some embodiments, thrombospondin type 1 domain variants comprise sequences substantially identical to any of the above-described sequences.

[116] To date, at least 1677 naturally occurring thrombospondin domains have been identified based on cDNA sequences. Exemplary proteins containing the naturally occurring thrombospondin domains include, *e.g.*, proteins in the complement pathway (*e.g.*, properdin, C6, C7, C8A, C8B, and C9), extracellular matrix proteins (*e.g.*, mindin, F-spondin, SCO-spondin,), circumsporozoite surface protein 2, and TRAP proteins of *Plasmodium*. Thrombospondin type 1 domains are further described in, *e.g.*, Roszmusz *et al.*, *BBRC* 296:156 (2002); Higgins *et al.*, *J Immunol.* 155:5777-85 (1995); Schultz-Cherry *et al.*, *J. Biol. Chem.* 270:7304-7310 (1995); Schultz-Cherry *et al.*, *J. Biol. Chem.* 269:26783-8 (1994); Bork, *FEBS Lett* 327:125-30 (1993); and Leung-Hagesteijn *et al.*, *Cell* 71:289-99 (1992).

[117] Another exemplary monomer domain suitable for use in the practice of the present invention is the trefoil domain. Trefoil monomer domains are typically about

about 30-50 or 30-65 amino acids. In some embodiments, the domains comprise about 35-55 amino acids and in some cases about 45 amino acids. Within the 35-55 amino acids, there are typically about 6 cysteine residues. Of the six cysteines, disulfide bonds typically are found between the following cysteines: C1 and C5, C2 and C4, C3 and C6.

[118] To date, at least 149 naturally occurring trefoil domains have identified based on cDNA sequences. Exemplary proteins containing naturally occurring trefoil domains include, *e.g.*, protein pS2 (TFF1), spasmolytic peptide SP (TFF2), intestinal trefoil factor (TFF3), intestinal surcease-isomaltase, and proteins which may be involved in defense against microbial infections by protecting the epithelia (*e.g.*, *Xenopus* xP1, xP4, integumentary mucins A.1 and C.1. Trefoil domains are further described in, *e.g.*, Sands and Podolsky, *Annu. Rev. Physiol.* 58:253-273 (1996); Carr *et al.*, *PNAS USA* 91:2206-2210 (1994); DeA *et al.*, *PNAS USA* 91:1084-1088 (1994); Hoffman *et al.*, *Trends Biochem Sci* 18:239-243 (1993).

[119] Exemplary trefoil domain sequences and consensus sequences are as follows:

(1) C_1(xx)xxxxxxxxxC_2xx(x)xxxxxxxC_3xxxxC_4C_5xxxxx(x)xxxxxC_6
(2) C_1(xx)xxxxxxRxxC_2xx(x)xxxxxxxC_3xxxxC_4C_5xxxxx(x)xxxxxC_6
(3) C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f
(4)
C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[qedk]C_3xx[krln][Gnk]C_4C_5[Fwy][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy]
(5)
C1(xx)xxx[Pvae]xxRx[ndpm]C2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C3xx[krln][Gnk]C4C5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C6[Fy]
(6)
C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C_6

[120] Another exemplary monomer domain suitable for use in the present invention is the thyroglobulin domain. Thyroglobulin monomer domains are typically about 30-85 or 30-80 amino acids. In some embodiments, the domains comprise about 35-75 amino acids and in some cases about 65 amino acids. Within the 35-75 amino acids, there are typically about 6 cysteine residues. Of the six cysteines, disulfide bonds typically are found between the following cysteines: C1 and C2, C3 and C4, C5 and C6.

[121] To date at least 251 naturally occurring thyroglobulin domains have been identified based on cDNA sequences. The N-terminal section of Tg contains 10 repeats of a domain of about 65 amino acids which is known as the Tg type-1 repeat

PUBMED:3595599, PUBMED:8797845. Exemplary proteins containing naturally occurring thyroglobulin domains include *e.g.*, the HLA class II associated invariant chain, human pancreatic carcinoma marker proteins, nidogen (entactin), insulin-like growth factor binding proteins (IGFBP), saxiphilin, chum salmon egg cysteine proteinase inhibitor, and equistatin. The Thyr-1 and related domains belong to MEROPS proteinase inhibitor family I31, clan IX. Thyroglobulin domains are further described in, *e.g.*, Molina *et al.*, *Eur. J. Biochem.* 240:125-133 (1996); Guncar *et al.*, *EMBO J* 18:793-803 (1999); Chong and Speicher, *DW* 276:5804-5813 (2001).

[122] Exemplary thyroglobulin domain sequences and consensus sequences are as follows:

(1)
C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxxxxxC_2xxxxxxxxxxC_3x(x)x(xxx)xxxxC_4x C_5xxxx(x)xxxxxxxxxxxxxx(xx)xC_6

(2)
C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxxxxxQC_3x(x)x(xxx)xxxxC_4 WC_5Vxxx(x)GxxxxGxxxxxxxx(xx)xC_6

(3)C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5 Vdxx(x)GxxxxGxxxxxgxx(xx)xC_6

(4)C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[s a]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6

(5)C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC2xxxGx[α]xx[vkrl]QC_3x(x[sa] xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[φg]xxxxxgxx(xx)xC_6

[123] Another exemplary monomer domain that can be used in the present invention is a laminin-EGF domain. Laminin-EGF domains are typically about 30-85 or 30-80 amino acids. In some embodiments, the domains comprise about 45-65 amino acids and in some cases about 50 amino acids. Within the 45-65 amino acids, there are typically about 8 cysteine residues which interact to form 4 disulfide bonds. Laminins are a major noncollagenous component of basement membranesthat mediate cell adhesion, growth migration, and differentiation. They are composed of distinct but related alpha, beta, and gamma chains. The three chains form a cross-shaped molecule that consist of a long arm and three short globular arms. The long arm consist of a coiled coil structure contributed by all three chains and cross-linked by interchain disulphide bonds.

[124] Exemplary laminin EGF domain sequences and consensus sequences are as follows:

(1)
C_1xC_2xxxxxx(xxx)xxC_3xxx(xxxxxx)xxxxC_4xC_5xxxxxxxxC_6xxC_7xxxxxxx(xxxxx)xxx xxC_8

(2)

$C_1xC_2xxxxxx(xxx)xxC3xxx(xxxxxx)xxgxC_4xC_5xxxxxGxxC_6xxC_7xxxxxxx(xxxxx)xxxxxC_8$

(3)

$C_1xC_2[ndh]xxxxx(xxx)xxC_3xxx(xxxxxx)xxgxC_4xC_5xxxxxGxxC_6[denq]xC_7xx[gn][yfht]xxx(xxxxx)xxxxxC_8$

[125] As mentioned above, monomer domains can be naturally-occurring or non-naturally occurring variants. The term "naturally occurring" is used herein to indicate that an object can be found in nature. For example, natural monomer domains can include human monomer domains or optionally, domains derived from different species or sources, *e.g.*, mammals, primates, rodents, fish, birds, reptiles, plants, *etc.* The natural occurring monomer domains can be obtained by a number of methods, *e.g.*, by PCR amplification of genomic DNA or cDNA. Libraries of monomer domains employed in the practice of the present invention may contain naturally-occurring monomer domain, non-naturally occurring monomer domain variants, or a combination thereof.

[126] Monomer domain variants can include ancestral domains, randomized domains, chimeric domains, mutated domains, and the like. For example, ancestral domains can be based on phylogenetic analysis. Randomized domains are domains in which one or more regions are randomized. The randomization can be based on full randomization, or optionally, partial randomization based on natural distribution of sequence diversity. Chimeric domains are domains in which one or more regions are replaced by corresponding regions from other domains of the same family. For example, chimeric domains can be constructed by combining loop sequences from multiple related domains of the same family to form novel domains with potentially lowered immunogenicity. Those of skill in the art will recognized the immunologic benefit of constructing modified binding domain monomers by combining loop regions from various related domains of the same family rather than creating random amino acid sequences. For example, by constructing variant domains by combining loop sequences or even multiple loop sequences that occur naturally in human thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains, the resulting domains may contain novel binding properties but may not contain any immunogenic protein sequences because all of the exposed loops are of human origin. The combining of loop amino acid sequences in endogenous context can be applied to all of the monomer constructs of the invention.

[127] The non-natural monomer domains or altered monomer domains can be produced by a number of methods. Any method of mutagenesis, such as site-directed mutagenesis and random mutagenesis (*e.g.*, chemical mutagenesis) can be used to produce

variants. In some embodiments, error-prone PCR is employed to create variants. Additional methods include aligning a plurality of naturally occurring monomer domains by aligning conserved amino acids in the plurality of naturally occurring monomer domains; and, designing the non-naturally occurring monomer domain by maintaining the conserved amino acids and inserting, deleting or altering amino acids around the conserved amino acids to generate the non-naturally occurring monomer domain. In one embodiment, the conserved amino acids comprise cysteines. In another embodiment, the inserting step uses random amino acids, or optionally, the inserting step uses portions of the naturally occurring monomer domains. The portions could ideally encode loops from domains from the same family. Amino acids are inserted or exchanged using synthetic oligonucleotides, or by shuffling, or by restriction enzyme based recombination. Human chimeric domains of the present invention are useful for therapeutic applications where minimal immunogenicity is desired. The present invention provides methods for generating libraries of human chimeric domains.

[128] Multimers or monomer domains of the invention can be produced according to any methods known in the art. In some embodiments, *E. coli* comprising a plasmid encoding the polypeptides under transcriptional control of a bacterial promoter are used to express the protein. After harvesting the bacteria, they may be lysed by sonication, heat, or homogenization and clarified by centrifugation. The polypeptides may be purified using Ni-NTA agarose elution (if 6xHis tagged) or DEAE sepharose elution (if untagged) and refolded by dialysis. Misfolded proteins may be neutralized by capping free sulfhydrils with iodoacetic acid. Q sepharose elution, butyl sepharose flow-through, SP sepharose elution, DEAE sepharose elution, and/or CM sepharose elution may be used to purify the polypeptides. Equivalent anion and/or cation exchange or hydrophobic interaction purification steps may also be employed.

[129] In some embodiments, monomers or multimers are purified using heat lysis, typically followed by a fast cooling to prevent most proteins from renaturing. Due to the heat stability of the proteins of the invention, the desired proteins will not be denatured by the heat and therefore will allow for a purification step (*i.e.*, purification that eliminates contaminant proteins) resulting in high purity. In some embodiments, a continuous flow heating process to purify the monomers or multimers from bacterial cell cultures is used. For example, a cell suspension can passed through a stainless steel coil submerged in a water bath set to a temperature resulting in lysis of the bacteria (*e.g.*, about 55°C, 60°C, 65°C, 70°C, 75°C, 80°C, 85°C, 90°C, 95°C, or 100°C for about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, or

60 minutes). The lysed effluent is routed to a cooling bath to obtain rapid cooling and prevent renaturation of denatured *E. coli* proteins. *E. coli* proteins denature and are prevented from renaturing, but the monomer or multimers do not denature under these conditions due to the exceptional stability of their scaffold. The heating time is controlled by adjusting the flow rate and length of the coil. This approach yields active proteins with high yield and exceptionally high purity (*e.g.*, >60%, >65%, >70%, >75%, or >80%) compared to alternative approaches and is amenable to high throughput (*e.g.*, 96-well or 384-well) production and large scale (*e.g.*, about 100 µl to about 1, 2, 5, 10, 15, 20, 50, 75, 100, 500, or 1000 liters) production of material including clinical material and material for screening assays (*e.g.*, *in vitro* binding and inhibition assays and cell-based activity assays) .

[130] In some embodiments, following manufacture of the monomers or multimers of the invention, the polypeptides are treated in a solution comprising iodoacetic acid to cap free -SH moieties of cysteines that have not formed disulfide bonds. In some embodiments, 0.1-100 mM (*e.g.*, 1-10 mM) iodoacetic acid is included in the solutions. Typically, the iodoacetic acid can be removed before administered to an individual.

[131] Polynucleotides (also referred to as nucleic acids) encoding the monomer domains are typically employed to make monomer domains via expression. Nucleic acids that encode monomer domains can be derived from a variety of different sources. Libraries of monomer domains can be prepared by expressing a plurality of different nucleic acids encoding naturally occurring monomer domains, altered monomer domains (*i.e.*, monomer domain variants), or a combinations thereof.

[132] Nucleic acids encoding fragments of naturally-occurring monomer domains and/or immuno-domains can also be mixed and/or recombined (*e.g.*, by using chemically or enzymatically-produced fragments) to generate full-length, modified monomer domains and/or immuno-domains. The fragments and the monomer domain can also be recombined by manipulating nucleic acids encoding domains or fragments thereof. For example, ligating a nucleic acid construct encoding fragments of the monomer domain can be used to generate an altered monomer domain.

[133] Altered monomer domains can also be generated by providing a collection of synthetic oligonucleotides (*e.g.*, overlapping oligonucleotides) encoding conserved, random, pseudorandom, or a defined sequence of peptide sequences that are then inserted by ligation into a predetermined site in a polynucleotide encoding a monomer domain. Similarly, the sequence diversity of one or more monomer domains can be expanded by mutating the monomer domain(s) with site-directed mutagenesis, random

mutation, pseudorandom mutation, defined kernal mutation, codon-based mutation, and the like. The resultant nucleic acid molecules can be propagated in a host for cloning and amplification. In some embodiments, the nucleic acids are recombined.

[134] The present invention also provides a method for recombining a plurality of nucleic acids encoding monomer domains and screening the resulting library for monomer domains that bind to the desired ligand or mixture of ligands or the like. Selected monomer domain nucleic acids can also be back-crossed by recombining with polynucleotide sequences encoding neutral sequences (*i.e.*, having insubstantial functional effect on binding), such as for example, by back-crossing with a wild-type or naturally-occurring sequence substantially identical to a selected sequence to produce native-like functional monomer domains. Generally, during back-crossing, subsequent selection is applied to retain the property, *e.g.*, binding to the ligand.

[135] In some embodiments, the monomer library is prepared by recombination. In such a case, monomer domains are isolated and recombined to combinatorially recombine the nucleic acid sequences that encode the monomer domains (recombination can occur between or within monomer domains, or both). The first step involves identifying a monomer domain having the desired property, *e.g.*, affinity for a certain ligand. While maintaining the conserved amino acids during the recombination, the nucleic acid sequences encoding the monomer domains can be recombined, or recombined and joined into multimers.

II. Multimers

[136] Methods for generating multimers (*i.e.*, recombinant mosaic proteins or combinatorial mosaic proteins) are a feature of the present invention. Multimers comprise at least two monomer domains . For example, multimers of the invention can comprise from 2 to about 10 monomer domains, from 2 and about 8 monomer domains, from about 3 and about 10 monomer domains, about 7 monomer domains, about 6 monomer domains, about 5 monomer domains, or about 4 monomer domains. In some embodiments, the multimer comprises at least 3 monomer domains. In view of the possible range of monomer domain sizes, the multimers of the invention may be, *e.g.*, 100 kD, 90kD, 80kD, 70kD, 60kD, 50kd, 40kD, 30kD, 25kD, 20kD, 15kD, 10kD, 5kD or smaller or larger. Typically, the monomer domains have been pre-selected for binding to the target molecule of interest.

[137] In some embodiments, each monomer domain specifically binds to one target molecule. In some of these embodiments, each monomer binds to a different position (analogous to an epitope) on a target molecule. Multiple monomer domains and/or immuno-domains that bind to the same target molecule result in an avidity effect yielding improved avidity of the multimer for the target molecule compared to each individual monomer. In some embodiments, the multimer has an avidity of at least about 1.5, 2, 3, 4, 5, 10, 20, 50 or 100 or 1000 times the avidity of a monomer domain alone. Typically, the multimer has a K_d of less than about 10^{-15}, 10^{-14}, 10^{-13}, 10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}, or 10^{-8}. In some embodiments, at least one, two, three, four or more (including all) monomers of a multimer bind an ion such as calcium or another ion.

[138] In another embodiment, the multimer comprises monomer domains with specificities for different target molecules. For example, multimers of such diverse monomer domains can specifically bind different components of a viral replication system or different serotypes of a virus. In some embodiments, at least one monomer domain binds to a toxin and at least one monomer domain binds to a cell surface molecule, thereby acting as a mechanism to target the toxin. In some embodiments, at least two monomer domains and/or immuno-domains of the multimer bind to different target molecules in a target cell or tissue. Similarly, therapeutic molecules can be targeted to the cell or tissue by binding a therapeutic agent to a monomer of the multimer that also contains other monomer domains and/or immuno-domains having cell or tissue binding specificity. In some embodiments, the different monomers bind to different components of a signal transduction pathway, a metabolic pathway, or components of different metabolic pathways that exert the same additive or synergistic physiological or biological effect or effects.

[139] Multimers can comprise a variety of combinations of monomer domains. For example, in a single multimer, the selected monomer domains can be the same or identical, optionally, different or non-identical. In addition, the selected monomer domains can comprise various different monomer domains from the same monomer domain family, or various monomer domains from different domain families, or optionally, a combination of both.

[140] Multimers that are generated in the practice of the present invention may be any of the following:

(1) A homo-multimer (a multimer of the same domain, *i.e.*, A1-A1-A1-A1);

(2) A hetero-multimer of different domains of the same domain class, *e.g.*, A1-A2-A3-A4. For example, hetero-multimer include multimers where A1, A2, A3 and A4 are different

non-naturally occurring variants of a particular thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains, or where some of A1, A2, A3, and A4 are naturally-occurring variants of a thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain.

(3) A hetero-multimer of domains from different monomer domain classes, *e.g.*, A1-B2-A2-B1. For example, where A1 and A2 are two different monomer domains (either naturally occurring or non-naturally-occurring) from thrombospondin type I, and B1 and B2 are two different monomer domains (either naturally occurring or non-naturally occurring) from a thyroglobulin.

[141] Multimer libraries employed in the practice of the present invention may contain homo-multimers, hetero-multimers of different monomer domains (natural or non-natural) of the same monomer class, or hetero-multimers of monomer domains (natural or non-natural) from different monomer classes, or combinations thereof. Other exemplary multimers include, *e.g.*, trimers and higher level (*e.g.*, tetramers).

[142] Monomer domains, as described herein, are also readily employed in a immuno-domain-containing heteromultimer (*i.e.*, a multimer that has at least one immuno-domain variant and one monomer domain variant). Thus, multimers of the present invention may have at least one immuno-domain such as a minibody, a single-domain antibody, a single chain variable fragment (ScFv), or a Fab fragment; and at least one monomer domain, such as, for example, a Thrombospondin type I domain, a thyroglobulin type I repeat domain, a Trefoil (P-type) domain, an EGF-like domain (*e.g.*, a Laminin-type EGF-like domain), a Kringle-domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin Inhibitor domain, a Kazal-type serine protease inhibitor domain, a von Willebrand factor type C domain, an Anaphylatoxin-like domain, a CUB domain LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type 3 domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a F5/8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, an EF Hand domain, a Cadherin domain, an Annexin domain, a zinc finger domain, and a C2 domain, or variants thereof.

[143] Domains need not be selected before the domains are linked to form multimers. On the other hand, the domains can be selected for the ability to bind to a target molecule before being linked into multimers. Thus, for example, a multimer can comprise

two domains that bind to one target molecule and a third domain that binds to a second target molecule.

[144] Typically, multimers of the present invention are a single discrete polypeptide. Multimers of partial linker-domain-partial linker moieties are an association of multiple polypeptides, each corresponding to a partial linker-domain-partial linker moiety.

[145] Accordingly, the multimers of the present invention may have the following qualities: multivalent, multispecific, single chain, heat stable, extended serum and/or shelf half-life. Moreover, at least one, more than one or all of the monomer domains may bind an ion (*e.g.*, a metal ion or a calcium ion), atleast one, more than one or all monomer domains may be derived from thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains, at least one, more than one or all of the monomer domains may be non-naturally occurring, and/or at least one, more than one or all of the monomer domains may comprise 1, 2, 3, or 4 disulfide bonds per monomer domain. In some embodiments, the multimers comprise at least two (or at least three) monomer domains, wherein at least one monomer domain is a non-naturally occurring monomer domain and the monomer domains bind calcium. In some embodiments, the multimers comprise at least 4 monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein:

a. each monomer domain is between 30-100 amino acids and each of the monomer domains comprise at least one disulfide linkage; or

b. each monomer domain is between 30-100 amino acids and is derived from an extracellular protein; or

c. each monomer domain is between 30-100 amino acids and binds to a protein target.

[146] In some embodiments, the multimers comprise at least 4 monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein:

a. each monomer domain is between 35-100 amino acids; or

b. each domain comprises at least one disulfide bond and is derived from a human protein and/or an extracellular protein.

[147] In some embodiments, the multimers comprise at least two monomer domains, wherein at least one monomer domain is non-naturally occurring, and wherein each domain is:

a. 25-50 amino acids long and comprises at least one disulfide bond; or

b. 25-50 amino acids long and is derived from an extracellular protein; or

c. 25-50 amino acids and binds to a protein target; or

d. 35-50 amino acids long.

[148] In some embodiments, the multimers comprise at least two monomer domains, wherein at least one monomer domain is non-naturally-occurring and:

a. each monomer domain comprises at least one disulfide bond; or

b. at least one monomer domain is derived from an extracellular protein; or

c. at least one monomer domain binds to a target protein.

[149] In some embodiments, the multimers of the invention bind to the same or other multimers to form aggregates. Aggregation can be mediated, for example, by the presence of hydrophobic domains on two monomer domains and/or immuno-domains, resulting in the formation of non-covalent interactions between two monomer domains and/or immuno-domains. Alternatively, aggregation may be facilitated by one or more monomer domains in a multimer having binding specificity for a monomer domain in another multimer. Aggregates can also form due to the presence of affinity peptides on the monomer domains or multimers. Aggregates can contain more target molecule binding domains than a single multimer.

[150] Multimers with affinity for both a cell surface target and a second target may provide for increased avidity effects. In some cases, membrane fluidity can be more flexible than protein linkers in optimizing (by self-assembly) the spacing and valency of the interactions. In some cases, multimers will bind to two different targets, each on a different cell or one on a cell and another on a molecule with multiple binding sites.

III. Linkers

[151] The selected monomer domains may be joined by a linker to form a single chain multimer. For example, a linker is positioned between each separate discrete monomer domain in a multimer. Typically, immuno-domains are also linked to each other or to monomer domains via a linker moiety. Linker moieties that can be readily employed to link immuno-domain variants together are the same as those described for multimers of monomer domain variants. Exemplary linker moieties suitable for joining immuno-domain variants to other domains into multimers are described herein.

[152] Joining the selected monomer domains via a linker can be accomplished using a variety of techniques known in the art. For example, combinatorial assembly of polynucleotides encoding selected monomer domains can be achieved by restriction digestion and re-ligation, by PCR-based, self-priming overlap reactions, or other

recombinant methods. The linker can be attached to a monomer before the monomer is identified for its ability to bind to a target multimer or after the monomer has been selected for the ability to bind to a target multimer.

[153] The linker can be naturally-occurring, synthetic or a combination of both. For example, the synthetic linker can be a randomized linker, *e.g.*, both in sequence and size. In one aspect, the randomized linker can comprise a fully randomized sequence, or optionally, the randomized linker can be based on natural linker sequences. The linker can comprise, e.g,. a non-polypeptide moiety, a polynucleotide, a polypeptide or the like.

[154] A linker can be rigid, or alternatively, flexible, or a combination of both. Linker flexibility can be a function of the composition of both the linker and the monomer domains that the linker interacts with. The linker joins two selected monomer domain, and maintains the monomer domains as separate discrete monomer domains. The linker can allow the separate discrete monomer domains to cooperate yet maintain separate properties such as multiple separate binding sites for the same ligand in a multimer, or *e.g.*, multiple separate binding sites for different ligands in a multimer. In some cases, a disulfide bridge exists between two linked monomer domains or between a linker and a monomer domain. In some embodiments, the monmer domains and/or linkers comprise metal-binding centers.

[155] Choosing a suitable linker for a specific case where two or more monomer domains (*i.e.* polypeptide chains) are to be connected may depend on a variety of parameters including, *e.g.* the nature of the monomer domains, the structure and nature of the target to which the polypeptide multimer should bind and/or the stability of the peptide linker towards proteolysis and oxidation.

[156] The present invention provides methods for optimizing the choice of linker once the desired monomer domains/variants have been identified. Generally, libraries of multimers having a composition that is fixed with regard to monomer domain composition, but variable in linker composition and length, can be readily prepared and screened as described above.

[157] Typically, the linker polypeptide may predominantly include amino acid residues selected from Gly, Ser, Ala and Thr. For example, the peptide linker may contain at least 75% (calculated on the basis of the total number of residues present in the peptide linker), such as at least 80%, *e.g.* at least 85% or at least 90% of amino acid residues selected from Gly, Ser, Ala and Thr. The peptide linker may also consist of Gly, Ser, Ala and/or Thr residues only. The linker polypeptide should have a length, which is adequate to

link two monomer domains in such a way that they assume the correct conformation relative to one another so that they retain the desired activity, for example as antagonists of a given receptor.

[158] A suitable length for this purpose is a length of at least one and typically fewer than about 50 amino acid residues, such as 2-25 amino acid residues, 5-20 amino acid residues, 5-15 amino acid residues, 8-12 amino acid residues or 11 residues. Similarly, the polypeptide encoding a linker can range in size, *e.g.*, from about 2 to about 15 amino acids, from about 3 to about 15, from about 4 to about 12, about 10, about 8, or about 6 amino acids. In methods and compositions involving nucleic acids, such as DNA, RNA, or combinations of both, the polynucleotide containing the linker sequence can be, *e.g.*, between about 6 nucleotides and about 45 nucleotides, between about 9 nucleotides and about 45 nucleotides, between about 12 nucleotides and about 36 nucleotides, about 30 nucleotides, about 24 nucleotides, or about 18 nucleotides. Likewise, the amino acid residues selected for inclusion in the linker polypeptide should exhibit properties that do not interfere significantly with the activity or function of the polypeptide multimer. Thus, the peptide linker should on the whole not exhibit a charge which would be inconsistent with the activity or function of the polypeptide multimer, or interfere with internal folding, or form bonds or other interactions with amino acid residues in one or more of the monomer domains which would seriously impede the binding of the polypeptide multimer to the target in question.

[159] In another embodiment of the invention, the peptide linker is selected from a library where the amino acid residues in the peptide linker are randomized for a specific set of monomer domains in a particular polypeptide multimer. A flexible linker could be used to find suitable combinations of monomer domains, which is then optimized using this random library of variable linkers to obtain linkers with optimal length and geometry. The optimal linkers may contain the minimal number of amino acid residues of the right type that participate in the binding to the target and restrict the movement of the monomer domains relative to each other in the polypeptide multimer when not bound to the target.

[160] The use of naturally occurring as well as artificial peptide linkers to connect polypeptides into novel linked fusion polypeptides is well known in the literature (Hallewell *et al.* (1989), *J. Biol. Chem.* 264, 5260-5268; Alfthan *et al.* (1995), *Protein Eng.* 8, 725-731; Robinson & Sauer (1996), *Biochemistry* 35, 109-116; Khandekar *et al.* (1997), *J. Biol. Chem.* 272, 32190-32197; Fares *et al.* (1998), *Endocrinology* 139, 2459-2464; Smallshaw *et al.* (1999), *Protein Eng.* 12, 623-630; US 5,856,456).

[161] One example where the use of peptide linkers is widespread is for production of single-chain antibodies where the variable regions of a light chain (V_L) and a heavy chain (V_H) are joined through an artificial linker, and a large number of publications exist within this particular field. A widely used peptide linker is a 15mer consisting of three repeats of a Gly-Gly-Gly-Gly-Ser amino acid sequence ($(Gly_4Ser)_3$). Other linkers have been used, and phage display technology, as well as, selective infective phage technology has been used to diversify and select appropriate linker sequences (Tang *et al.* (1996), *J. Biol. Chem.* 271, 15682-15686; Hennecke *et al.* (1998), *Protein Eng.* 11, 405-410). Peptide linkers have been used to connect individual chains in hetero- and homo-dimeric proteins such as the T-cell receptor, the lambda Cro repressor, the P22 phage Arc repressor, IL-12, TSH, FSH, IL-5, and interferon-γ. Peptide linkers have also been used to create fusion polypeptides. Various linkers have been used and in the case of the Arc repressor phage display has been used to optimize the linker length and composition for increased stability of the single-chain protein (Robinson and Sauer (1998), *Proc. Natl. Acad. Sci. USA* 95, 5929-5934).

[162] Another type of linker is an intein, *i.e.* a peptide stretch which is expressed with the single-chain polypeptide, but removed post-translationally by protein splicing. The use of inteins is reviewed by F.S. Gimble in *Chemistry and Biology*, 1998, Vol 5, No. 10 pp. 251-256.

[163] Still another way of obtaining a suitable linker is by optimizing a simple linker, *e.g.* $(Gly_4Ser)_n$, through random mutagenesis.

[164] As mentioned above, it is generally preferred that the peptide linker possess at least some flexibility. Accordingly, in some embodiments, the peptide linker contains 1-25 glycine residues, 5-20 glycine residues, 5-15 glycine residues or 8-12 glycine residues. The peptide linker will typically contain at least 50% glycine residues, such as at least 75% glycine residues. In some embodiments of the invention, the peptide linker comprises glycine residues only.

[165] The peptide linker may, in addition to the glycine residues, comprise other residues, in particular residues selected from Ser, Ala and Thr, in particular Ser. Thus, one example of a specific peptide linker includes a peptide linker having the amino acid sequence Gly_x-Xaa-Gly_y-Xaa-Gly_z, wherein each Xaa is independently selected from the group consisting Ala, Val, Leu, Ile, Met, Phe, Trp, Pro, Gly, Ser, Thr, Cys, Tyr, Asn, Gln, Lys, Arg, His, Asp and Glu, and wherein x, y and z are each integers in the range from 1-5. In some embodiments, each Xaa is independently selected from Ser, Ala and Thr, in

particular Ser. More particularly, the peptide linker has the amino acid sequence Gly-Gly-Gly-Xaa-Gly-Gly-Gly-Xaa-Gly-Gly-Gly, wherein each Xaa is independently selected from the group consisting Ala, Val, Leu, Ile, Met, Phe, Trp, Pro, Gly, Ser, Thr, Cys, Tyr, Asn, Gln, Lys, Arg, His, Asp and Glu. In some embodiments, each Xaa is independently selected from Ser, Ala and Thr, in particular Ser.

[166] In some cases it may be desirable or necessary to provide some rigidity into the peptide linker. This may be accomplished by including proline residues in the amino acid sequence of the peptide linker. Thus, in another embodiment of the invention, the peptide linker comprises at least one proline residue in the amino acid sequence of the peptide linker. For example, the peptide linker has an amino acid sequence, wherein at least 25%, such as at least 50%, *e.g.* at least 75%, of the amino acid residues are proline residues. In one particular embodiment of the invention, the peptide linker comprises proline residues only.

[167] In some embodiments of the invention, the peptide linker is modified in such a way that an amino acid residue comprising an attachment group for a non-polypeptide moiety is introduced. Examples of such amino acid residues may be a cysteine residue (to which the non-polypeptide moiety is then subsequently attached) or the amino acid sequence may include an *in vivo* N-glycosylation site (thereby attaching a sugar moiety (*in vivo*) to the peptide linker). An additional option is to genetically incorporate non-natural amino acids using evolved tRNAs and tRNA synthetases (*see, e.g.*, U.S. Patent Application Publication 2003/0082575) into the monomer domains or linkers. For example, insertion of keto-tyrosine allows for site-specific coupling to expressed monomer domains or multimers.

[168] In some embodiments of the invention, the peptide linker comprises at least one cysteine residue, such as one cysteine residue. Thus, in some embodiments of the invention the peptide linker comprises amino acid residues selected from Gly, Ser, Ala, Thr and Cys. In some embodiments, such a peptide linker comprises one cysteine residue only.

[169] In a further embodiment, the peptide linker comprises glycine residues and cysteine residue, such as glycine residues and cysteine residues only. Typically, only one cysteine residue will be included per peptide linker. Thus, one example of a specific peptide linker comprising a cysteine residue, includes a peptide linker having the amino acid sequence Gly_n-Cys-Gly_m, wherein n and m are each integers from 1-12, *e.g.*, from 3-9, from 4-8, or from 4-7. More particularly, the peptide linker may have the amino acid sequence GGGGG-C-GGGGG.

[170] This approach (*i.e.* introduction of an amino acid residue comprising an attachment group for a non-polypeptide moiety) may also be used for the more rigid proline-containing linkers. Accordingly, the peptide linker may comprise proline and cysteine residues, such as proline and cysteine residues only. An example of a specific proline-containing peptide linker comprising a cysteine residue, includes a peptide linker having the amino acid sequence Pro_n-Cys-Pro_m, wherein n and m are each integers from 1-12, preferably from 3-9, such as from 4-8 or from 4-7. More particularly, the peptide linker may have the amino acid sequence PPPPP-C-PPPPP.

[171] In some embodiments, the purpose of introducing an amino acid residue, such as a cysteine residue, comprising an attachment group for a non-polypeptide moiety is to subsequently attach a non-polypeptide moiety to said residue. For example, non-polypeptide moieties can improve the serum half-life of the polypeptide multimer. Thus, the cysteine residue can be covalently attached to a non-polypeptide moiety. Preferred examples of non-polypeptide moieties include polymer molecules, such as PEG or mPEG, in particular mPEG as well as non-polypeptide therapeutic agents.

[172] The skilled person will acknowledge that amino acid residues other than cysteine may be used for attaching a non-polypeptide to the peptide linker. One particular example of such other residue includes coupling the non-polypeptide moiety to a lysine residue.

[173] Another possibility of introducing a site-specific attachment group for a non-polypeptide moiety in the peptide linker is to introduce an *in vivo* N-glycosylation site, such as one *in vivo* N-glycosylation site, in the peptide linker. For example, an *in vivo* N-glycosylation site may be introduced in a peptide linker comprising amino acid residues selected from Gly, Ser, Ala and Thr. It will be understood that in order to ensure that a sugar moiety is in fact attached to said *in vivo* N-glycosylation site, the nucleotide sequence encoding the polypeptide multimer must be inserted in a glycosylating, eukaryotic expression host.

[174] A specific example of a peptide linker comprising an *in vivo* N-glycosylation site is a peptide linker having the amino acid sequence Gly_n-Asn-Xaa-Ser/Thr-Gly_m, preferably Gly_n-Asn-Xaa-Thr-Gly_m, wherein Xaa is any amino acid residue except proline, and wherein n and m are each integers in the range from 1-8, preferably in the range from 2-5.

[175] Often, the amino acid sequences of all peptide linkers present in the polypeptide multimer will be identical. Nevertheless, in certain embodiments the amino acid

sequences of all peptide linkers present in the polypeptide multimer may be different. The latter is believed to be particular relevant in case the polypeptide multimer is a polypeptide tri-mer or tetra-mer and particularly in such cases where an amino acid residue comprising an attachment group for a non-polypeptide moiety is included in the peptide linker.

[176] Quite often, it will be desirable or necessary to attach only a few, typically only one, non-polypeptide moieties/moiety (such as mPEG, a sugar moiety or a non-polypeptide therapeutic agent) to the polypeptide multimer in order to achieve the desired effect, such as prolonged serum-half life. Evidently, in case of a polypeptide tri-mer, which will contain two peptide linkers, only one peptide linker is typically required to be modified, *e.g.* by introduction of a cysteine residue, whereas modification of the other peptide linker will typically not be necessary not. In this case all (both) peptide linkers of the polypeptide multimer (tri-mer) are different.

[177] Accordingly, in a further embodiment of the invention, the amino acid sequences of all peptide linkers present in the polypeptide multimer are identical except for one, two or three peptide linkers, such as except for one or two peptide linkers, in particular except for one peptide linker, which has/have an amino acid sequence comprising an amino acid residue comprising an attachment group for a non-polypeptide moiety. Preferred examples of such amino acid residues include cysteine residues of *in vivo* N-glycosylation sites.

[178] A linker can be a native or synthetic linker sequence. An exemplary native linker includes, *e.g.*, the sequence between the last cysteine of a first thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain and the first cysteine of a second thrombospondin type I monomer domain, thyroglobulin monomer domain, or trefoil monomer domain can be used as a linker sequence. Analysis of various domain linkages reveals that native linkers range from at least 3 amino acids to fewer than 20 amino acids, *e.g.*, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 amino acids long. However, those of skill in the art will recognize that longer or shorter linker sequences can be used. In some embodiments, the linker is a 6-mer of the following sequence $A_1A_2A_3A_4A_5A_6$, wherein A_1 is selected from the amino acids A, P, T, Q, E and K; A_2 and A_3 are any amino acid except C, F, Y, W, or M; A_4 is selected from the amino acids S, G and R; A_5 is selected from the amino acids H, P, and R; and A_6 is the amino acid, T.

[179] Methods for generating multimers from monomer domains and/or immuno-domains can include joining the selected domains with at least one linker to generate at least one multimer, *e.g.*, the multimer can comprise at least two of the monomer domains

and/or immuno-domains and the linker. The multimer(s) is then screened for an improved avidity or affinity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains. A composition of the multimer produced by the method is included in the present invention.

[180] In other methods, the selected multimer domains are joined with at least one linker to generate at least two multimers, wherein the two multimers comprise two or more of the selected monomer domains and the linker. The two or more multimers are screened for an improved avidity or affinity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains. Compositions of two or more multimers produced by the above method are also features of the invention.

[181] Linkers, multimers or selected multimers produced by the methods indicated above and below are features of the present invention. Libraries comprising multimers, e.g, a library comprising about 100, 250, 500 or more members produced by the methods of the present invention or selected by the methods of the present invention are provided. In some embodiments, one or more cell comprising members of the libraries, are also included. Libraries of the recombinant polypeptides are also a feature of the present invention, *e.g.*, a library comprising about 100, 250, 500 or more different recombinant polypetides.

[182] Suitable linkers employed in the practice of the present invention include an obligate heterodimer of partial linker moieties. The term "obligate heterodimer" (also referred to as "affinity peptides") refers herein to a dimer of two partial linker moieties that differ from each other in composition, and which associate with each other in a non-covalent, specific manner to join two domains together. The specific association is such that the two partial linkers associate substantially with each other as compared to associating with other partial linkers. Thus, in contrast to multimers of the present invention that are expressed as a single polypeptide, multimers of domains that are linked together via heterodimers are assembled from discrete partial linker-monomer-partial linker units. Assembly of the heterodimers can be achieved by, for example, mixing. Thus, if the partial linkers are polypeptide segments, each partial linker-monomer-partial linker unit may be expressed as a discrete peptide prior to multimer assembly. A disulfide bond can be added to covalently lock the peptides together following the correct non-covalent pairing. Partial linker moieties that are appropriate for forming obligate heterodimers include, for example, polynucleotides, polypeptides, and the like. For example, when the partial linker is a polypeptide, binding domains are produced individually along with their unique linking

peptide (*i.e.*, a partial linker) and later combined to form multimers. *See, e.g.*, Madden, M., Aldwin, L., Gallop, M. A., and Stemmer, W. P. C. (1993) Peptide linkers: Unique self-associative high-affinity peptide linkers. Thirteenth American Peptide Symposium, Edmonton, Canada (abstract). The spatial order of the binding domains in the multimer is thus mandated by the heterodimeric binding specificity of each partial linker. Partial linkers can contain terminal amino acid sequences that specifically bind to a defined heterologous amino acid sequence. An example of such an amino acid sequence is the Hydra neuropeptide head activator as described in Bodenmuller *et al.*, *The neuropeptide head activator loses its biological activity by dimerization*, (1986) EMBO J 5(8):1825-1829. *See, e.g.*, U.S. Patent No. 5,491,074 and WO 94/28173. These partial linkers allow the multimer to be produced first as monomer-partial linker units or partial linker-monomer-partial linker units that are then mixed together and allowed to assemble into the ideal order based on the binding specificities of each partial linker. Alternatively, monomers linked to partial linkers can be contacted to a surface, such as a cell, in which multiple monomers can associate to form higher avidity complexes via partial linkers. In some cases, the association will form via random Brownian motion.

[183] When the partial linker comprises a DNA binding motif, each monomer domain has an upstream and a downstream partial linker (*i.e.*, Lp-domain-Lp, where "Lp" is a representation of a partial linker) that contains a DNA binding protein with exclusively unique DNA binding specificity. These domains can be produced individually and then assembled into a specific multimer by the mixing of the domains with DNA fragments containing the proper nucleotide sequences (*i.e.*, the specific recognition sites for the DNA binding proteins of the partial linkers of the two desired domains) so as to join the domains in the desired order. Additionally, the same domains may be assembled into many different multimers by the addition of DNA sequences containing various combinations of DNA binding protein recognition sites. Further randomization of the combinations of DNA binding protein recognition sites in the DNA fragments can allow the assembly of libraries of multimers. The DNA can be synthesized with backbone analogs to prevent degradation in vivo.

[184] In some embodiments, the multimer comprises monomer domains with specificities for different proteins. The different proteins can be related or unrelated. Examples of related proteins including members of a protein family or different serotypes of a virus. Alternatively, the monomer domains of a multimer can target different molecules in a physiological pathway (*e.g.*, different blood coagulation proteins). In yet other

embodiments, monomer domains bind to proteins in unrelated pathways (*e.g.*, two domains bind to blood factors, two other domains bind to inflammation-related proteins and a fifth binds to serum albumin). In another embodiment, a multimer is comprised of monomer domains that bind to different pathogens or contaminants of interest. Such multimers are useful to as a single detection agent capable of detecting for the possibility of any of a number of pathogens or contaminants.

IV. Methods of Identifying Monomer Domains and/or Multimers with a Desired Binding Affinity

[185] The invention provides methods of identifying monomer domains that bind to a selected or desired ligand or mixture of ligands. In some embodiments, monomer domains and/or immuno-domains are identified or selected for a desired property (*e.g.*, binding affinity) and then the monomer domains and/or immuno-domains are formed into multimers. For those embodiments, any method resulting in selection of domains with a desired property (*e.g.*, a specific binding property) can be used. For example, the methods can comprise providing a plurality of different nucleic acids, each nucleic acid encoding a monomer domain; translating the plurality of different nucleic acids, thereby providing a plurality of different monomer domains; screening the plurality of different monomer domains for binding of the desired ligand or a mixture of ligands; and, identifying members of the plurality of different monomer domains that bind the desired ligand or mixture of ligands.

[186] Selection of monomer domains and/or immuno-domains from a library of domains can be accomplished by a variety of procedures. For example, one method of identifying monomer domains and/or immuno-domains which have a desired property involves translating a plurality of nucleic acids, where each nucleic acid encodes a monomer domain and/or immuno-domain, screening the polypeptides encoded by the plurality of nucleic acids, and identifying those monomer domains and/or immuno-domains that, *e.g.*, bind to a desired ligand or mixture of ligands, thereby producing a selected monomer domain and/or immuno-domain. The monomer domains and/or immuno-domains expressed by each of the nucleic acids can be tested for their ability to bind to the ligand by methods known in the art (*i.e.* panning, affinity chromatography, FACS analysis).

[187] As mentioned above, selection of monomer domains and/or immuno-domains can be based on binding to a ligand such as a target protein or other target molecule

(*e.g.*, lipid, carbohydrate, nucleic acid and the like). Other molecules can optionally be included in the methods along with the target, *e.g.*, ions such as Ca^{+2}. The ligand can be a known ligand, *e.g.*, a ligand known to bind one of the plurality of monomer domains, or *e.g.*, the desired ligand can be an unknown monomer domain ligand. Other selections of monomer domains and/or immuno-domains can be based, *e.g.*, on inhibiting or enhancing a specific function of a target protein or an activity. Target protein activity can include, *e.g.*, endocytosis or internalization, induction of second messenger system, up-regulation or down-regulation of a gene, binding to an extracellular matrix, release of a molecule(s), or a change in conformation. In this case, the ligand does not need to be known. The selection can also include using high-throughput assays.

[188] When a monomer domain and/or immuno-domain is selected based on its ability to bind to a ligand, the selection basis can include selection based on a slow dissociation rate, which is usually predictive of high affinity. The valency of the ligand can also be varied to control the average binding affinity of selected monomer domains and/or immuno-domains. The ligand can be bound to a surface or substrate at varying densities, such as by including a competitor compound, by dilution, or by other method known to those in the art. High density (valency) of predetermined ligand can be used to enrich for monomer domains that have relatively low affinity, whereas a low density (valency) can preferentially enrich for higher affinity monomer domains.

[189] A variety of reporting display vectors or systems can be used to express nucleic acids encoding the monomer domains immuno-domains and/or multimers of the present invention and to test for a desired activity. For example, a phage display system is a system in which monomer domains are expressed as fusion proteins on the phage surface (Pharmacia, Milwaukee Wis.). Phage display can involve the presentation of a polypeptide sequence encoding monomer domains and/or immuno-domains on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein.

[190] Generally in these methods, each phage particle or cell serves as an individual library member displaying a single species of displayed polypeptide in addition to the natural phage or cell protein sequences. The plurality of nucleic acids are cloned into the phage DNA at a site which results in the transcription of a fusion protein, a portion of which is encoded by the plurality of the nucleic acids. The phage containing a nucleic acid molecule undergoes replication and transcription in the cell. The leader sequence of the fusion protein directs the transport of the fusion protein to the tip of the phage particle. Thus, the fusion protein that is partially encoded by the nucleic acid is displayed on the phage

particle for detection and selection by the methods described above and below. For example, the phage library can be incubated with a predetermined (desired) ligand, so that phage particles which present a fusion protein sequence that binds to the ligand can be differentially partitioned from those that do not present polypeptide sequences that bind to the predetermined ligand. For example, the separation can be provided by immobilizing the predetermined ligand. The phage particles (*i.e.*, library members) which are bound to the immobilized ligand are then recovered and replicated to amplify the selected phage subpopulation for a subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the phage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed polypeptide sequence is determined, thereby identifying the sequence(s) of polypeptides that bind to the predetermined ligand. Such methods are further described in PCT patent publication Nos. 91/17271, 91/18980, and 91/19818 and 93/08278.

[191] Examples of other display systems include ribosome displays, a nucleotide-linked display (*see*, *e.g.*, U.S. Patent Nos. 6,281,344; 6,194,550, 6,207,446, 6,214,553, and 6,258,558), polysome display, cell surface displays and the like. The cell surface displays include a variety of cells, *e.g.*, E. coli, yeast and/or mammalian cells. When a cell is used as a display, the nucleic acids, *e.g.*, obtained by PCR amplification followed by digestion, are introduced into the cell and translated. Optionally, polypeptides encoding the monomer domains or the multimers of the present invention can be introduced, *e.g.*, by injection, into the cell.

[192] Those of skill in the art will recognize that the steps of generating variation and screening for a desired property can be repeated (*i.e.*, performed recursively) to optimize results. For example, in a phage display library or other like format, a first screening of a library can be performed at relatively lower stringency, thereby selected as many particles associated with a target molecule as possible. The selected particles can then be isolated and the polynucleotides encoding the monomer or multimer can be isolated from the particles. Additional variations can then be generated from these sequences and subsequently screened at higher affinity.

[193] Monomer domains may be selected to bind any type of target molecule, including protein targets. Exemplary targets include, but are not limited to, *e.g.*, IL-6, Alpha3, cMet, ICOS, IgE, IL-1-R11, BAFF, CD40L, CD28, Her2, TRAIL-R, VEGF,

TPO-R, TNFα, LFA-1, TACI, IL-1b, B7.1, B7.2, or OX40. When the target is a receptor for a ligand, the monomer domains may act as antagonists or agonists of the receptor.

[194] When multimers capable of binding relatively large targets are desired, they can be generated by a "walking" selection method. As shown in Figure 3, this method is carried out by providing a library of monomer domains and screening the library of monomer domains for affinity to a first target molecule. Once at least one monomer that binds to the target is identified, that particular monomer is covalently linked to a new library or each remaining member of the original library of monomer domains. The new library members each comprise one common domain and at least one domain that that is different, *i.e.*, randomized. Thus, in some embodiments, the invention provides a library of multimers generated using the "walking" selection method. This new library of multimers (*e.g.*, dimers, trimers, tetramers, and the like) is then screened for multimers that bind to the target with an increased affinity, and a multimer that binds to the target with an increased affinity can be identified. The "walking" monomer selection method provides a way to assemble a multimer that is composed of monomers that can act additively or even synergistically with each other given the restraints of linker length. This walking technique is very useful when selecting for and assembling multimers that are able to bind large target proteins with high affinity. The walking method can be repeated to add more monomers thereby resulting in a multimer comprising 2, 3, 4, 5, 6, 7, 8 or more monomers linked together.

[195] In some embodiments, the selected multimer comprises more than two domains. Such multimers can be generated in a step fashion, *e.g.*, where the addition of each new domain is tested individually and the effect of the domains is tested in a sequential fashion. In an alternate embodiment, domains are linked to form multimers comprising more than two domains and selected for binding without prior knowledge of how smaller multimers, or alternatively, how each domain, bind.

[196] The methods of the present invention also include methods of evolving monomers or multimers. As illustrated in Figure 10, intra-domain recombination can be introduced into monomers across the entire monomer or by taking portions of different monomers to form new recombined units. The different monomers may bind the same target or different targets. For example, in some embodiments portions of different thrombospondin monomers may be recombined. In some embdiments, a portion of a thrombospondin monomer may be combined with a portion of a thyroglobulin monomer and/or a portion of a trefoil/PD monomer. Interdomain recombination (*e.g.*, recombining different monomers into

or between multimers) or recombination of modules (*e.g.*, multiple monomers within a multimer) may be achieved. Inter-library recombination is also contemplated.

[197] Figure 8 illustrates the process of intradomain optimization by recombination. Shown is a three-fragment PCR overlap reaction, which recombines three segments of a single domain relative to each other. One can use two, three, four, five or more fragment overlap reactions in the same way as illustrated. This recombination process has many applications. One application is to recombine a large pool of hundreds of previously selected clones without sequence information. All that is needed for each overlap to work is one known region of (relatively) constant sequence that exists in the same location in each of the clones (fixed site approach). The intra-domain recombination method can also be performed on a pool of sequence-related monomer domains by standard DNA recombination (*e.g.*, Stemmer, *Nature* 370:389-391 (1994)) based on random fragmentation and reassembly based on DNA sequence homology, which does not require a fixed overlap site in all of the clones that are to be recombined.

[198] Another application of this process is to create multiple separate, naïve (meaning unpanned) libraries in each of which only one of the intercysteine loops is randomized, to randomize a different loop in each library. After panning of these libraries separately against the target, the selected clones are then recombined. From each panned library only the randomized segment is amplified by PCR and multiple randomized segments are then combined into a single domain, creating a shuffled library which is panned and/or screened for increased potency. This process can also be used to shuffle a small number of clones of known sequence.

[199] Any common sequence may be used as cross-over points. For cysteine-containing monomers, the cysteine residues are logical places for the crossover. However, there are other ways to determine optimal crossover sites, such as computer modeling. Alternatively, residues with highest entropy, or the least number of intramolecular contacts, may also be good sites for crossovers.

[200] Methods for evolving monomers or multimers can comprise, *e.g.*, any or all of the following steps: providing a plurality of different nucleic acids, where each nucleic acid encoding a monomer domain; translating the plurality of different nucleic acids, which provides a plurality of different monomer domains; screening the plurality of different monomer domains for binding of the desired ligand or mixture of ligands; identifying members of the plurality of different monomer domains that bind the desired ligand or mixture of ligands, which provides selected monomer domains; joining the selected monomer

domains with at least one linker to generate at least one multimer, wherein the at least one multimer comprises at least two of the selected monomer domains and the at least one linker; and, screening the at least one multimer for an improved affinity or avidity or altered specificity for the desired ligand or mixture of ligands as compared to the selected monomer domains.

[201] Variation can be introduced into either monomers or multimers. As discussed above, an example of improving monomers includes intra-domain recombination in which two or more (*e.g.*, three, four, five, or more) portions of the monomer are amplified separately under conditions to introduce variation (for example by shuffling or other recombination method) in the resulting amplification products, thereby synthesizing a library of variants for different portions of the monomer. By locating the 5' ends of the middle primers in a "middle" or 'overlap' sequence that both of the PCR fragments have in common, the resulting "left" side and "right" side libraries may be combined by overlap PCR to generate novel variants of the original pool of monomers. These new variants may then be screened for desired properties, *e.g.*, panned against a target or screened for a functional effect. The "middle" primer(s) may be selected to correspond to any segment of the monomer, and will typically be based on the scaffold or one or more concensus amino acids within the monomer (*e.g.*, cysteines such as those found in A domains).

[202] Similarly, multimers may be created by introducing variation at the monomer level and then recombining monomer variant libraries. On a larger scale, multimers (single or pools) with desired properties may be recombined to form longer multimers. In some cases variation is introduced (typically synthetically) into the monomers or into the linkers to form libraries. This may be achieved, *e.g.*, with two different multimers that bind to two different targets, thereby eventually selecting a multimer with a portion that binds to one target and a portion that binds a second target. *See, e.g.*, Figure 9.

[203] Additional variation can be introduced by inserting linkers of different length and composition between domains. This allows for the selection of optimal linkers between domains. In some embodiments, optimal length and composition of linkers will allow for optimal binding of domains. In some embodiments, the domains with a particular binding affinity(s) are linked via different linkers and optimal linkers are selected in a binding assay. For example, domains are selected for desired binding properties and then formed into a library comprising a variety of linkers. The library can then be screened to identify optimal linkers. Alternatively, multimer libraries can be formed where the effect of domain or linker on target molecule binding is not known.

[204] Methods of the present invention also include generating one or more selected multimers by providing a plurality of monomer domains and/or immuno-domains. The plurality of monomer domains and/or immuno-domains is screened for binding of a desired ligand or mixture of ligands. Members of the plurality of domains that bind the desired ligand or mixture of ligands are identified, thereby providing domains with a desired affinity. The identified domains are joined with at least one linker to generate the multimers, wherein each multimer comprises at least two of the selected domains and the at least one linker; and, the multimers are screened for an improved affinity or avidity or altered specificity for the desired ligand or mixture of ligands as compared to the selected domains, thereby identifying the one or more selected multimers.

[205] Multimer libraries may be generated, in some embodiments, by combining two or more libraries or monomers or multimers in a recombinase-based approach, where each library member comprises as recombination site (*e.g.*, a lox site). A larger pool of molecularly diverse library members in principle harbor more variants with desired properties, such as higher target-binding affinities and functional activities. When libraries are constructed in phage vectors, which may be transformed into *E. coli*, library size (10^9 - 10^{10}) is limited by the transformation efficiency of *E. coli*. A recombinase/recombination site system (*e.g.*, the *Cre-lox*P system) and *in vivo* recombination can be exploited to generate libraries that are not limited in size by the transformation efficiency of *E. coli*.

[206] For example, the Cre-loxP system may be used to generate dimer libraries with 10^{10}, 10^{11}, 10^{12}, 10^{13}, or greater diversity. In some embodiments, *E. coli* as a host for one naïve monomer library and a filamentous phage that carries a second naïve monomer library are used. The library size in this case is limited only by the number of infective phage (carrying one library) and the number of infectible *E. coli* cells (carrying the other library). For example, infecting 10^{12} *E. coli* cells (1L at OD600=1) with $>10^{12}$ phage could produce as many as 10^{12} dimer combinations.

[207] Selection of multimers can be accomplished using a variety of techniques including those mentioned above for identifying monomer domains. Other selection methods include, *e.g.*, a selection based on an improved affinity or avidity or altered specificity for the ligand compared to selected monomer domains. For example, a selection can be based on selective binding to specific cell types, or to a set of related cells or protein types (*e.g.*, different virus serotypes). Optimization of the property selected for, *e.g.*, avidity of a ligand, can then be achieved by recombining the domains, as well as manipulating amino

acid sequence of the individual monomer domains or the linker domain or the nucleotide sequence encoding such domains, as mentioned in the present invention.

[208] One method for identifying multimers can be accomplished by displaying the multimers. As with the monomer domains, the multimers are optionally expressed or displayed on a variety of display systems, *e.g.*, phage display, ribosome display, polysome display, nucleotide-linked display (*see, e.g.*, U.S. Patent Nos. 6,281,344; 6,194,550, 6,207,446, 6,214,553, and 6,258,558) and/or cell surface display, as described above. Cell surface displays can include but are not limited to *E. coli*, yeast or mammalian cells. In addition, display libraries of multimers with multiple binding sites can be panned for avidity or affinity or altered specificity for a ligand or for multiple ligands.

[209] Monomers or multimers can be screened for target binding activity in yeast cells using a two-hybrid screening assay. In this type of screen the monomer or multimer library to be screened is cloned into a vector that directs the formation of a fusion protein between each monomer or multimer of the library and a yeast transcriptional activator fragment (*i.e.*, Gal4). Sequences encoding the "target" protein are cloned into a vector that results in the production of a fusion protein between the target and the remainder of the Gal4 protein (the DNA binding domain). A third plasmid contains a reporter gene downstream of the DNA sequence of the Gal4 binding site. A monomer that can bind to the target protein brings with it the Gal4 activation domain, thus reconstituting a functional Gal4 protein. This functional Gal4 protein bound to the binding site upstream of the reporter gene results in the expression of the reporter gene and selection of the monomer or multimer as a target binding protein. (see Chien et.al. (1991) *Proc. Natl. Acad. Sci.* (USA) 88:9578; Fields S. and Song O. (1989) *Nature* 340: 245) Using a two-hybrid system for library screening is further described in U.S. Patent No. 5,811,238 (see also Silver S.C. and Hunt S.W. (1993) *Mol. Biol. Rep.* 17:155; Durfee *et al.* (1993) *Genes Devel.* 7:555; Yang *et al.* (1992) *Science* 257:680; Luban *et al.* (1993) *Cell* 73:1067; Hardy *et al.* (1992) *Genes Devel.* 6:801; Bartel *et al.* (1993) *Biotechniques* 14:920; and Vojtek *et al.* (1993) *Cell* 74:205). Another useful screening system for carrying out the present invention is the *E.coli*/BCCP interactive screening system (Germino *et al.* (1993) *Proc. Nat. Acad. Sci.* (U.S.A.) 90:993; Guarente L. (1993) *Proc. Nat. Acad. Sci.* (U.S.A.) 90:1639).

[210] Other variations include the use of multiple binding compounds, such that monomer domains, multimers or libraries of these molecules can be simultaneously screened for a multiplicity of ligands or compounds that have different binding specificity. Multiple predetermined ligands or compounds can be concomitantly screened in a single

library, or sequential screening against a number of monomer domains or multimers. In one variation, multiple ligands or compounds, each encoded on a separate bead (or subset of beads), can be mixed and incubated with monomer domains, multimers or libraries of these molecules under suitable binding conditions. The collection of beads, comprising multiple ligands or compounds, can then be used to isolate, by affinity selection, selected monomer domains, selected multimers or library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two individual ligands or compounds. This approach affords efficient screening, and is compatible with laboratory automation, batch processing, and high throughput screening methods.

[211] In another embodiment, multimers can be simultaneously screened for the ability to bind multiple ligands, wherein each ligand comprises a different label. For example, each ligand can be labeled with a different fluorescent label, contacted simultaneously with a multimer or multimer library. Multimers with the desired affinity are then identified (*e.g.*, by FACS sorting) based on the presence of the labels linked to the desired labels.

[212] Libraries of either monomer domains or multimers (referred in the following discussion for convenience as "affinity agents") can be screened (*i.e.*, panned) simultaneously against multiple ligands in a number of different formats. For example, multiple ligands can be screened in a simple mixture, in an array, displayed on a cell or tissue (*e.g.*, a cell or tissue provides numerous molecules that can be bound by the monomer domains or multimers of the invention), and/or immobilized. *See*, *e.g.*, Figure 4. The libraries of affinity agents can optionally be displayed on yeast or phage display systems. Similarly, if desired, the ligands (*e.g.*, encoded in a cDNA library) can be displayed in a yeast or phage display system.

[213] Initially, the affinity agent library is panned against the multiple ligands. Optionally, the resulting "hits" are panned against the ligands one or more times to enrich the resulting population of affinity agents.

[214] If desired, the identity of the individual affinity agents and/or ligands can be determined. In some embodiments, affinity agents are displayed on phage. Affinity agents identified as binding in the initial screen are divided into a first and second portion. The first portion is infected into bacteria, resulting in either plaques or bacterial colonies,

depending on the type of phage used. The expressed phage are immobilized and then probed with ligands displayed in phage selected as described below.

[215] The second portion are coupled to beads or otherwise immobilized and a phage display library containing at least some of the ligands in the original mixture is contacted to the immobilized second portion. Those phage that bind to the second portion are subsequently eluted and contacted to the immobilized phage described in the paragraph above. Phage-phage interactions are detected (*e.g.*, using a monoclonal antibody specific for the ligand-expressing phage) and the resulting phage polynucleotides can be isolated.

[216] In some embodiments, the identity of an affinity agent-ligand pair is determined. For example, when both the affinity agent and the ligand are displayed on a phage or yeast, the DNA from the pair can be isolated and sequenced. In some embodiments, polynucleotides specific for the ligand and affinity agent are amplified. Amplification primers for each reaction can include 5' sequences that are complementary such that the resulting amplification products are fused, thereby forming a hybrid polynucleotide comprising a polynucleotide encoding at least a portion of the affinity agent and at least a portion of the ligand. The resulting hybrid can be used to probe affinity agent or ligand (*e.g.*, cDNA-encoded) polynucleotide libraries to identify both affinity agent and ligand. *See, e.g.*, Figure 10.

[217] The above-described methods can be readily combined with "walking" to simultaneous generate and identify multiple multimers, each of which bind to a ligand in a mixture of ligands. In these embodiments, a first library of affinity agents (monomer domains, immuno domains or multimers) are panned against multiple ligands and the eluted affinity agents are linked to the first or a second library of affinity agents to form a library of multimeric affinity agents (*e.g.*, comprising 2, 3, 4, 5, 6, 7, 8, 9, or more monomer or immuno domains), which are subsequently panned against the multiple ligands. This method can be repeated to continue to generate larger multimeric affinity agents. Increasing the number of monomer domains may result in increased affinity and avidity for a particular target. Of course, at each stage, the panning is optionally repeated to enrich for significant binders. In some cases, walking will be facilitated by inserting recombination sites (*e.g.*, lox sites) at the ends of monomers and recombining monomer libraries by a recombinase-mediated event.

[218] The selected multimers of the above methods can be further manipulated, *e.g.*, by recombining or shuffling the selected multimers (recombination can occur between or within multimers or both), mutating the selected multimers, and the like. This results in altered multimers which then can be screened and selected for members that

have an enhanced property compared to the selected multimer, thereby producing selected altered multimers.

[219] In view of the description herein, it is clear that the following process may be followed. Naturally or non-naturally occurring monomer domains may be recombined or variants may be formed. Optionally the domains initially or later are selected for those sequences that are less likely to be immunogenic in the host for which they are intended. Optionally, a phage library comprising the recombined domains is panned for a desired affinity. Monomer domains or multimers expressed by the phage may be screened for IC_{50} for a target. Hetero- or homo-meric multimers may be selected. The selected polypeptides may be selected for their affinity to any target, including, *e.g.*, hetero- or homo-multimeric targets.

[220] A significant advantage of the present invention is that known ligands, or unknown ligands can be used to select the monomer domains and/or multimers. No prior information regarding ligand structure is required to isolate the monomer domains of interest or the multimers of interest. The monomer domains and/or multimers identified can have biological activity, which is meant to include at least specific binding affinity for a selected or desired ligand, and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like. Monomer domains can be generated to function as ligands for receptors where the natural ligand for the receptor has not yet been identified (orphan receptors). These orphan ligands can be created to either block or activate the receptor top which they bind.

[221] A single ligand can be used, or optionally a variety of ligands can be used to select the monomer domains and/or multimers. A monomer domain and/or immuno-domain of the present invention can bind a single ligand or a variety of ligands. A multimer of the present invention can have multiple discrete binding sites for a single ligand, or optionally, can have multiple binding sites for a variety of ligands.

V. Libraries

[222] The present invention also provides libraries of monomer domains and libraries of nucleic acids that encode monomer domains and/or immuno-domains. The libraries can include, *e.g.*, about 10, 100, 250, 500, 1000, or 10,000 or more nucleic acids encoding monomer domains, or the library can include, *e.g.*, about 10, 100, 250, 500, 1000 or

10,000 or more polypeptides that encode monomer domains. Libraries can include monomer domains containing the same cysteine frame, *e.g.*, thrombosponding domains, thyroglobulin domains, or trefoil/PD domains.

[223] In some embodiments, variants are generated by recombining two or more different sequences from the same family of monomer domains (*e.g.*, the LDL receptor class A domain). Alternatively, two or more different monomer domains from different families can be combined to form a multimer. In some embodiments, the multimers are formed from monomers or monomer variants of at least one of the following family classes: a thrombospondin type I domain, a thyroglobulin type I repeat domain, a Trefoil (P-type) domain, an EGF-like domain (*e.g.*, a Laminin-type EGF-like domain), a Kringle-domain, a fibronectin type I domain, a fibronectin type II domain, a fibronectin type III domain, a PAN domain, a Gla domain, a SRCR domain, a Kunitz/Bovine pancreatic trypsin Inhibitor domain, a Kazal-type serine protease inhibitor domain, a von Willebrand factor type C domain, an Anaphylatoxin-like domain, a CUB domain LDL-receptor class A domain, a Sushi domain, a Link domain, a Thrombospondin type 3 domain, an Immunoglobulin-like domain, a C-type lectin domain, a MAM domain, a von Willebrand factor type A domain, a Somatomedin B domain, a WAP-type four disulfide core domain, a F5/8 type C domain, a Hemopexin domain, an SH2 domain, an SH3 domain, an EF Hand domain, a Cadherin domain, an Annexin domain, a zinc finger domain, and a C2 domain and derivatives thereof. In another embodiment, the monomer domain and the different monomer domain can include one or more domains found in the Pfam database and/or the SMART database. Libraries produced by the methods above, one or more cell(s) comprising one or more members of the library, and one or more displays comprising one or more members of the library are also included in the present invention.

[224] Optionally, a data set of nucleic acid character strings encoding monomer domains can be generated *e.g.*, by mixing a first character string encoding a monomer domain, with one or more character string encoding a different monomer domain, thereby producing a data set of nucleic acids character strings encoding monomer domains, including those described herein. In another embodiment, the monomer domain and the different monomer domain can include one or more domains found in the Pfam database and/or the SMART database. The methods can further comprise inserting the first character string encoding the monomer domain and the one or more second character string encoding the different monomer domain in a computer and generating a multimer character string(s) or library(s), thereof in the computer.

[225] The libraries can be screened for a desired property such as binding of a desired ligand or mixture of ligands or otherwise exposed to selective conditions. For example, members of the library of monomer domains can be displayed and prescreened for binding to a known or unknown ligand or a mixture of ligands or incubated in serum to remove those clones that are sensitive to serum proteases. The monomer domain sequences can then be mutagenized (*e.g.*, recombined, chemically altered, *etc.*) or otherwise altered and the new monomer domains can be screened again for binding to the ligand or the mixture of ligands with an improved affinity. The selected monomer domains can be combined or joined to form multimers, which can then be screened for an improved affinity or avidity or altered specificity for the ligand or the mixture of ligands. Altered specificity can mean that the specificity is broadened, *e.g.*, binding of multiple related viruses, or optionally, altered specificity can mean that the specificity is narrowed, *e.g.*, binding within a specific region of a ligand. Those of skill in the art will recognize that there are a number of methods available to calculate avidity. *See, e.g.*, Mammen *et al.*, *Angew Chem Int. Ed.* 37:2754-2794 (1998); Muller *et al.*, *Anal. Biochem.* 261:149-158 (1998).

[226] The present invention also provides a method for generating a library of chimeric monomer domains derived from human proteins, the method comprising: providing loop sequences corresponding to at least one loop from each of at least two different naturally occurring variants of a human protein, wherein the loop sequences are polynucleotide or polypeptide sequences; and covalently combining loop sequences to generate a library of at least two different chimeric sequences, wherein each chimeric sequence encodes a chimeric monomer domain having at least two loops. Typically, the chimeric domain has at least four loops, and usually at least six loops. As described above, the present invention provides three types of loops that are identified by specific features, such as, potential for disulfide bonding, bridging between secondary protein structures, and molecular dynamics (*i.e.*, flexibility). The three types of loop sequences are a cysteine-defined loop sequence, a structure-defined loop sequence, and a B-factor-defined loop sequence.

[227] Alternatively, a human chimeric domain library can be generated by modifying naturally occurring human monomer domains at the amino acid level, as compared to the loop level. To minimize the potential for immunogenicity, only those residues that naturally occur in protein sequences from the same family of human monomer domains are utilized to create the chimeric sequences. This can be achieved by providing a sequence alignment of at least two human monomer domains from the same family of monomer

domains, identifying amino acid residues in corresponding positions in the human monomer domain sequences that differ between the human monomer domains, generating two or more human chimeric monomer domains, wherein each human chimeric monomer domain sequence consists of amino acid residues that correspond in type and position to residues from two or more human monomer domains from the same family of monomer domains. Libraries of human chimeric monomer domains can be employed to identify human chimeric monomer domains that bind to a target of interest by: screening the library of human chimeric monomer domains for binding to a target molecule, and identifying a human chimeric monomer domain that binds to the target molecule. Suitable naturally occurring human monomer domain sequences employed in the initial sequence alignment step include those corresponding to any of the naturally occurring monomer domains described herein.

[228] Human chimeric domain libraries of the present invention (whether generated by varying loops or single amino acid residues) can be prepared by methods known to those having ordinary skill in the art. Methods particularly suitable for generating these libraries are split-pool format and trinucleotide synthesis format as described in WO01/23401.

VI. Fusion Proteins

[229] In some embodiments, the monomers or multimers of the present invention are linked to another polypeptide to form a fusion protein. Any polypeptide in the art may be used as a fusion partner, though it can be useful if the fusion partner forms multimers. For example, monomers or multimers of the invention may, for example, be fused to the following locations or combinations of locations of an antibody:

1. At the N-terminus of the VH1 and/or VL1 domains, optionally just after the leader peptide and before the domain starts (framework region 1);
2. At the N-terminus of the CH1 or CL1 domain, replacing the VH1 or VL1 domain;
3. At the N-terminus of the heavy chain, optionally after the CH1 domain and before the cysteine residues in the hinge (Fc-fusion);
4. At the N-terminus of the CH3 domain;
5. At the C-terminus of the CH3 domain, optionally attached to the last amino acid residue via a short linker;
6. At the C-terminus of the CH2 domain, replacing the CH3 domain;

7. At the C-terminus of the CL1 or CH1 domain, optionally after the cysteine that forms the interchain disulfide; or

8. At the C-terminus of the VH1 or VL1 domain. *See, e.g.*, Figure 7.

[230] In some embodiments, the monomer or multimer domain is linked to a molecule (*e.g.*, a protein, nucleic acid, organic small molecule, *etc.*) useful as a pharmaceutical. Exemplary pharmaceutical proteins include, *e.g.*, cytokines, antibodies, chemokines, growth factors, interleukins, cell-surface proteins, extracellular domains, cell surface receptors, cytotoxins, *etc.* Exemplary small molecule pharmaceuticals include small molecule toxins or therapeutic agents.

[231] In some embodiments, the monomer or multimers are selected to bind to a tissue- or disease-specific target protein. Tissue-specific proteins are proteins that are expressed exclusively, or at a significantly higher level, in one or several particular tissue(s) compared to other tissues in an animal. Similarly, disease-specific proteins are proteins that are expressed exclusively, or at a significantly higher level, in one or several diseased cells or tissues compared to other non-diseased cells or tissues in an animal. Examples of such diseases include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathycandidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's

syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and GerstmannStraussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular

acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. Exemplary disease or conditions include, *e.g.*, MS, SLE, ITP, IDDM, MG, CLL, CD, RA, Factor VIII Hemophilia, transplantation, arteriosclerosis, Sjogren's Syndrome, Kawasaki Disease, anti-phospholipid Ab, AHA, ulcerative colitis, multiple myeloma, Glomerulonephritis, seasonal allergies, and IgA Nephropathy.

[232] In some embodiments, the monomers or multimers that bind to the target protein are linked to the pharmaceutical protein or small molecule such that the resulting complex or fusion is targeted to the specific tissue or disease-related cell(s) where the target protein is expressed. Monomers or multimers for use in such complexes or fusions can be initially selected for binding to the target protein and may be subsequently selected by negative selection against other cells or tissue (*e.g.*, to avoid targeting bone marrow or other tissues that set the lower limit of drug toxicity) where it is desired that binding be reduced or eliminated in other non-target cells or tissues. By keeping the pharmaceutical away from sensitive tissues, the therapeutic window is increased so that a higher dose may be administered safely. In another alternative, *in vivo* panning can be performed in animals by injecting a library of monomers or multimers into an animal and then isolating the monomers or multimers that bind to a particular tissue or cell of interest.

[233] The fusion proteins described above may also include a linker peptide between the pharmaceutical protein and the monomer or multimers. A peptide linker sequence may be employed to separate, for example, the polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Fusion proteins can also be expressed as recombinant proteins in an expression system by standard techniques.

[234] Exemplary tissue-specific or disease-specific proteins can be found in, *e.g.*, Tables I and II of U.S. Patent Publication No 2002/0107215. Exemplary tissues where target proteins may be specifically expressed include, *e.g.*, liver, pancreas, adrenal gland, thyroid, salivary gland, pituitary gland, brain, spinal cord, lung, heart, breast, skeletal

muscle, bone marrow, thymus, spleen, lymph node, colorectal, stomach, ovarian, small intestine, uterus, placenta, prostate, testis, colon, colon, gastric, bladder, trachea, kidney, or adipose tissue.

VII. Compositions

[235] The invention also includes compositions that are produced by methods of the present invention. For example, the present invention includes monomer domains selected or identified from a library and/or libraries comprising monomer domains produced by the methods of the present invention.

[236] Compositions of nucleic acids and polypeptides are included in the present invention. For example, the present invention provides a plurality of different nucleic acids wherein each nucleic acid encodes at least one monomer domain or immuno-domain. In some embodiments, at least one monomer domain is selected from: an EGF-like domain (*e.g.*, a laminin EGF domain), a Trefoil (P-type) domain, a thyroglobulin type I repeat, a Thrombospondin type I domain, and variants of one or more thereof. Suitable monomer domains also include those listed in the Pfam database and/or the SMART database.

[237] The present invention also provides recombinant nucleic acids encoding one or more polypeptides comprising a plurality of monomer domains, which monomer domains are altered in order or sequence as compared to a naturally occuring polypeptide. For example, the naturally occuring polypeptide can be selected from: an EGF-like domain (*e.g.*, a laminin EGF domain), a Trefoil (P-type) domain, a thyroglobulin type I repeat domain, a Thrombospondin type I domain, and variants of one or more thereof. In another embodiment, the naturally occuring polypeptide encodes a monomer domain found in the Pfam database and/or the SMART database.

[238] All the compositions of the present invention, including the compositions produced by the methods of the present invention, *e.g.*, monomer domains as well as multimers and libraries thereof can be optionally bound to a matrix of an affinity material. Examples of affinity material include beads, a column, a solid support, a microarray, other pools of reagent-supports, and the like. In some embodiments, screening in solution uses a target that has been biotinylated. In these embodiments, the target is incubated with the phage library and the targets with the bound phage, are captured using streptavidin beads.

[239] Compositions of the present invention can be bound to a matrix of an affinity material, *e.g.*, the recombinant polypeptides. Examples of affinity material include, *e.g.*, beads, a column, a solid support, and/or the like.

VIII. Therapeutic and Prophylactic Treatment Methods

[240] The present invention also includes methods of therapeutically or prophylactically treating a disease or disorder by administering *in vivo* or *ex vivo* one or more nucleic acids or polypeptides of the invention described above (or compositions comprising a pharmaceutically acceptable excipient and one or more such nucleic acids or polypeptides) to a subject, including, *e.g.*, a mammal, including a human, primate, mouse, pig, cow, goat, rabbit, rat, guinea pig, hamster, horse, sheep; or a non-mammalian vertebrate such as a bird (*e.g.*, a chicken or duck), fish, or invertebrate.

[241] In one aspect of the invention, in *ex vivo* methods, one or more cells or a population of cells of interest of the subject (*e.g.*, tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosae, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, *etc.*) are obtained or removed from the subject and contacted with an amount of a selected monomer domain and/or multimer of the invention that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition. The contacted cells are then returned or delivered to the subject to the site from which they were obtained or to another site (*e.g.*, including those defined above) of interest in the subject to be treated. If desired, the contacted cells can be grafted onto a tissue, organ, or system site (including all described above) of interest in the subject using standard and well-known grafting techniques or, *e.g.*, delivered to the blood or lymph system using standard delivery or transfusion techniques.

[242] The invention also provides *in vivo* methods in which one or more cells or a population of cells of interest of the subject are contacted directly or indirectly with an amount of a selected monomer domain and/or multimer of the invention effective in prophylactically or therapeutically treating the disease, disorder, or other condition. In direct contact/administration formats, the selected monomer domain and/or multimer is typically administered or transferred directly to the cells to be treated or to the tissue site of interest (*e.g.*, tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosae, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, *etc.*) by any of a variety of formats, including topical administration,

injection (*e.g.*, by using a needle or syringe), or vaccine or gene gun delivery, pushing into a tissue, organ, or skin site. The selected monomer domain and/or multimer can be delivered, for example, intramuscularly, intradermally, subdermally, subcutaneously, orally, intraperitoneally, intrathecally, intravenously, or placed within a cavity of the body (including, *e.g.*, during surgery), or by inhalation or vaginal or rectal administration. In some embodiments, the proteins of the invention are prepared at concentrations of at least 25 mg/ml, 50 mg/ml, 75 mg/ml, 100 mg/ml, 150 mg/ml or more. Such concentrations are useful, for example, for subcutaneous formulations.

[243] In *in vivo* indirect contact/administration formats, the selected monomer domain and/or multimer is typically administered or transferred indirectly to the cells to be treated or to the tissue site of interest, including those described above (such as, *e.g.*, skin cells, organ systems, lymphatic system, or blood cell system, *etc.*), by contacting or administering the polypeptide of the invention directly to one or more cells or population of cells from which treatment can be facilitated. For example, tumor cells within the body of the subject can be treated by contacting cells of the blood or lymphatic system, skin, or an organ with a sufficient amount of the selected monomer domain and/or multimer such that delivery of the selected monomer domain and/or multimer to the site of interest (*e.g.*, tissue, organ, or cells of interest or blood or lymphatic system within the body) occurs and effective prophylactic or therapeutic treatment results. Such contact, administration, or transfer is typically made by using one or more of the routes or modes of administration described above.

[244] In another aspect, the invention provides *ex vivo* methods in which one or more cells of interest or a population of cells of interest of the subject (*e.g.*, tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosae, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, *etc.*) are obtained or removed from the subject and transformed by contacting said one or more cells or population of cells with a polynucleotide construct comprising a nucleic acid sequence of the invention that encodes a biologically active polypeptide of interest (*e.g.*, a selected monomer domain and/or multimer) that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition. The one or more cells or population of cells is contacted with a sufficient amount of the polynucleotide construct and a promoter controlling expression of said nucleic acid sequence such that uptake of the polynucleotide construct (and promoter) into the cell(s) occurs and sufficient expression of the target nucleic acid sequence of the invention results to produce an amount of the

biologically active polypeptide, encoding a selected monomer domain and/or multimer, effective to prophylactically or therapeutically treat the disease, disorder, or condition. The polynucleotide construct can include a promoter sequence (*e.g.*, CMV promoter sequence) that controls expression of the nucleic acid sequence of the invention and/or, if desired, one or more additional nucleotide sequences encoding at least one or more of another polypeptide of the invention, a cytokine, adjuvant, or co-stimulatory molecule, or other polypeptide of interest.

[245] Following transfection, the transformed cells are returned, delivered, or transferred to the subject to the tissue site or system from which they were obtained or to another site (*e.g.*, tumor cells, tumor tissue sample, organ cells, blood cells, cells of the skin, lung, heart, muscle, brain, mucosae, liver, intestine, spleen, stomach, lymphatic system, cervix, vagina, prostate, mouth, tongue, *etc.*) to be treated in the subject. If desired, the cells can be grafted onto a tissue, skin, organ, or body system of interest in the subject using standard and well-known grafting techniques or delivered to the blood or lymphatic system using standard delivery or transfusion techniques. Such delivery, administration, or transfer of transformed cells is typically made by using one or more of the routes or modes of administration described above. Expression of the target nucleic acid occurs naturally or can be induced (as described in greater detail below) and an amount of the encoded polypeptide is expressed sufficient and effective to treat the disease or condition at the site or tissue system.

[246] In another aspect, the invention provides *in vivo* methods in which one or more cells of interest or a population of cells of the subject (*e.g.*, including those cells and cells systems and subjects described above) are transformed in the body of the subject by contacting the cell(s) or population of cells with (or administering or transferring to the cell(s) or population of cells using one or more of the routes or modes of administration described above) a polynucleotide construct comprising a nucleic acid sequence of the invention that encodes a biologically active polypeptide of interest (*e.g.*, a selected monomer domain and/or multimer) that is effective in prophylactically or therapeutically treating the disease, disorder, or other condition.

[247] The polynucleotide construct can be directly administered or transferred to cell(s) suffering from the disease or disorder (*e.g.*, by direct contact using one or more of the routes or modes of administration described above). Alternatively, the polynucleotide construct can be indirectly administered or transferred to cell(s) suffering from the disease or disorder by first directly contacting non-diseased cell(s) or other diseased cells using one or more of the routes or modes of administration described above with a

sufficient amount of the polynucleotide construct comprising the nucleic acid sequence encoding the biologically active polypeptide, and a promoter controlling expression of the nucleic acid sequence, such that uptake of the polynucleotide construct (and promoter) into the cell(s) occurs and sufficient expression of the nucleic acid sequence of the invention results to produce an amount of the biologically active polypeptide effective to prophylactically or therapeutically treat the disease or disorder, and whereby the polynucleotide construct or the resulting expressed polypeptide is transferred naturally or automatically from the initial delivery site, system, tissue or organ of the subject's body to the diseased site, tissue, organ or system of the subject's body (*e.g.*, via the blood or lymphatic system). Expression of the target nucleic acid occurs naturally or can be induced (as described in greater detail below) such that an amount of expressed polypeptide is sufficient and effective to treat the disease or condition at the site or tissue system. The polynucleotide construct can include a promoter sequence (*e.g.*, CMV promoter sequence) that controls expression of the nucleic acid sequence and/or, if desired, one or more additional nucleotide sequences encoding at least one or more of another polypeptide of the invention, a cytokine, adjuvant, or co-stimulatory molecule, or other polypeptide of interest.

[248] In each of the *in vivo* and *ex vivo* treatment methods as described above, a composition comprising an excipient and the polypeptide or nucleic acid of the invention can be administered or delivered. In one aspect, a composition comprising a pharmaceutically acceptable excipient and a polypeptide or nucleic acid of the invention is administered or delivered to the subject as described above in an amount effective to treat the disease or disorder.

[249] In another aspect, in each *in vivo* and *ex vivo* treatment method described above, the amount of polynucleotide administered to the cell(s) or subject can be an amount such that uptake of said polynucleotide into one or more cells of the subject occurs and sufficient expression of said nucleic acid sequence results to produce an amount of a biologically active polypeptide effective to enhance an immune response in the subject, including an immune response induced by an immunogen (*e.g.*, antigen). In another aspect, for each such method, the amount of polypeptide administered to cell(s) or subject can be an amount sufficient to enhance an immune response in the subject, including that induced by an immunogen (*e.g.*, antigen).

[250] In yet another aspect, in an in vivo or in vivo treatment method in which a polynucleotide construct (or composition comprising a polynucleotide construct) is used to deliver a physiologically active polypeptide to a subject, the expression of the

polynucleotide construct can be induced by using an inducible on- and off-gene expression system. Examples of such on- and off-gene expression systems include the Tet-On™ Gene Expression System and Tet-Off™ Gene Expression System (*see, e.g.*, Clontech Catalog 2000, pg. 110-111 for a detailed description of each such system), respectively. Other controllable or inducible on- and off-gene expression systems are known to those of ordinary skill in the art. With such system, expression of the target nucleic of the polynucleotide construct can be regulated in a precise, reversible, and quantitative manner. Gene expression of the target nucleic acid can be induced, for example, after the stable transfected cells containing the polynucleotide construct comprising the target nucleic acid are delivered or transferred to or made to contact the tissue site, organ or system of interest. Such systems are of particular benefit in treatment methods and formats in which it is advantageous to delay or precisely control expression of the target nucleic acid (*e.g.*, to allow time for completion of surgery and/or healing following surgery; to allow time for the polynucleotide construct comprising the target nucleic acid to reach the site, cells, system, or tissue to be treated; to allow time for the graft containing cells transformed with the construct to become incorporated into the tissue or organ onto or into which it has been spliced or attached, *etc.*).

IX. Additional Multimer Uses

[251] The potential applications of multimers of the present invention are diverse and include any use where an affinity agent is desired. For example, the invention can be used in the application for creating antagonists, where the selected monomer domains or multimers block the interaction between two proteins. Optionally, the invention can generate agonists. For example, multimers binding two different proteins, *e.g.*, enzyme and substrate, can enhance protein function, including, for example, enzymatic activity and/or substrate conversion.

[252] Other applications include cell targeting. For example, multimers consisting of monomer domains and/or immuno-domains that recognize specific cell surface proteins can bind selectively to certain cell types. Applications involving monomer domains and/or immuno-domains as antiviral agents are also included. For example, multimers binding to different epitopes on the virus particle can be useful as antiviral agents because of the polyvalency. Other applications can include, but are not limited to, protein purification, protein detection, biosensors, ligand-affinity capture experiments and the like. Furthermore,

domains or multimers can be synthesized in bulk by conventional means for any suitable use, *e.g.*, as a therapeutic or diagnostic agent.

[253] In some embodiments, the monomer domains are used for ligand inhibition, ligand clearance or ligand stimulation. Possible ligands in these methods, include, *e.g.*, cytokines, chemokines, or growth factors.

[254] If inhibition of ligand binding to a receptor is desired, a monomer domain is selected that binds to the ligand at a portion of the ligand that contacts the ligand's receptor, or that binds to the receptor at a portion of the receptor that binds contacts the ligand, thereby preventing the ligand-receptor interaction. The monomer domains can optionally be linked to a half-life extender, if desired.

[255] Ligand clearance refers to modulating the half-life of a soluble ligand in bodily fluid. For example, most monomer domains, absent a half-life extender, have a short half-life. Thus, binding of a monomer domain to the ligand will reduce the half-life of the ligand, thereby reducing ligand concentration. The portion of the ligand bound by the monomer domain will generally not matter, though it may be beneficial to bind the ligand at the portion of the ligand that binds to its receptor, thereby further inhibiting the ligand's effect. This method is useful for reducing the concentration of any molecule in the bloodstream. In some embodiments, the concentration of a molecule in the bloodstream is reduced by enhancing the rate of kidney clearance of the molecule. Typically the monomer domain-molecule complex is less than about 40 KDa, less than about 50 KDa, or less than about 60 KDa.

[256] Alternatively, a multimer comprising a first monomer domain that binds to a half-life extender and a second monomer domain that binds to a portion of the ligand that does not bind to the ligand's receptor can be used to increase the half-life of the ligand.

[257] The invention further provide monomer domains that bind to a blood factor (*e.g.*, serum albumin, immunoglobulin, or erythrocytes).

[258] In some embodiments, the the monomer domains bind to an immunoglobulin polypeptide or a portion thereof.

[259] Four families (*i.e.*, Families 1, 2, 3 and 4) of monomer domains that bind to immunoglobulin have been identified.

[260] Sequences for Family 1 are set forth below. Dashes are included only for spacing.

Fam1

CASGQFQCRSTSICVPMWWRCDGVPDCPDNSDEK--SCEPP----T-------
CASGQFQCRSTSICVPMWWRCDGVPDCVDNSDET--SCTST----VHT-----
CASGQFQCRSTSICVPMWWRCDGVPDCADGSDEK--DCQQH----T-------
CASGQFQCRSTSICVPMWWRCDGVNDCGDGSDEA--DCGRPGPGATSAPAA--
CASGQFQCRSTSICVPMWWRCDGVPDCLDSSDEK--SCNAP----ASEPPGSL
CASGQFQCRSTSICVPMWWRCDGVPDCRDGSDEAPAHCSAP----ASEPPGSL
CASGQFQCRSTSICVPQWWVCDGVPDCRDGSDEP-EQCTPP----T-------
CLSSQFRCRDTGICVPQWWVCDGVPDCGDGSDEKG--CGRT----GHT-----
CLSSQFRCRDTGICVPQWWVCDGVPDCRDGSDEAAV-CGRP----GHT-----
CLSSQFRCRDTGICVPQWWVCDGVPDCRDGSDEAPAHCSAP----ASEPPGSL

[261] Family 2 has the following motif:

[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXXDCXD[DN]SDE

[262] Exemplary sequences comprising the IgG Family 2 motif are set forht below. Dashes are included only for spacing.

Fam2

CGAS-EFTCRSSSRCIPQAWVCDGENDCRDNSDE--ADCSAPASEPPGSL
CRSN-EFTCRSSERCIPLAWVCDGDNDCRDDSDE--ANCSAPASEPPGSL
CVSN-EFQCRGTRRCIPRTWLCDGLPDCGDNSDEAPANCSAPASEPPGSL
CHPTGQFRCRSSGRCVSPTWVCDGDNDCGDNSDE--ENCSAPASEPPGSL
CQAG-EFQC-GNGRCISPAWVCDGENDCRDGSDE--ANCSAPASEPPGSL

[263] Family 3 has either of the two following motifs:

CXSSGRCIPXXWVCDGXXDCRDXSDE; or
CXSSGRCIPXXWLCDGXXDCRDXSDE

[264] Exemplary sequences comprising the IgG Family 3 motif are set forth below. Dashes are included only for spacing.

Fam3

CPPSQFTCKSNDKCIPVHWLCDGDNDCGDSSDE--ANCGRPGPGATSAPAA
CPSGEFPCRSSGRCIPLAWLCDGDNDCRDNSDEPPALCGRPGPGATSAPAA
CAPSEFQCRSSGRCIPLPWVCDGEDDCRDGSDES-AVCGAPAP--T-----
CQASEFTCKSSGRCIPQEWLCDGEDDCRDSSDE--KNCQQPT---------
CLSSEFQCQSSGRCIPLAWVCDGDNDCRDDSDE--KSCKPRT---------

[265] Based on family 3 alignments, additional non-naturally occurring monomer domains that bind IgG and that has the sequence SSGR immediately preceding the third cysteine in an A domain scaffold. The sequences of these monomer domains are set forth below. Dashes are included only for spacing.

Fam4

CPANEFQCSNGRCISPAWLCDGENDCVDGSDE--KGCTPRT
CPPSEFQCGNGRCISPAWLCDGDNDCVDGSDE--TNCTTSGPT
CPPGEFQCGNGRCISAGWVCDGENDCVDDSDE--KDCPART
CGSGEFQCSNGRCISLGWVCDGEDDCPDGSDE--TNCGDSHILPFSTPGPST
CPADEFTCGNGRCISPAWVCDGEPDCRDGSDE-AAVCETHT
CPSNEFTCGNGRCISLAWLCDGEPDCRDSSDESLAICSQDPEFHKV

[266] Monomer domains that bind to red blood cells (RBC) or serum albumin (CSA) are described in U.S. Patent Publication No. 2005/0048512, and include, e.g.,:

```
RBCA    CRSSQFQCNDSRICIPGRWRCDGDNDCQDGSDETGCGDSHILPFSTPGPST
RBCB    CPAGEFPCKNGQCLPVTWLCDGVNDCLDGSDEKGCGRPGPGATSAPAA
RBC11   CPPDEFPCKNGQCIPQDWLCDGVNDCLDGSDEKDCGRPGPGATSAPAA
CSA-A8 CGAGQFPCKNGHCLPLNLLCDGVNDCEDNSDEPSELCKALT
```

[267] The present invention provides a method for extending the serum half-life of a protein, including, *e.g.*, a multimer of the invention or a protein of interest in an animal. The protein of interest can be any protein with therapeutic, prophylactic, or otherwise desirable functionality (including another monomer domain or multimer of the present invention). This method comprises first providing a monomer domain that has been identified as a binding protein that specifically binds to a half-life extender such as a blood-carried molecule or cell, such as serum proteins such as albumin (*e.g.*, human serum albumin) or transferrin, IgG or a portion thereof, red blood cells, *etc.* In some embodiments, the half-life extender-binding monomer can be covalently linked to another monomer domain that has a binding affinity for the protein of interest. This multimer, optionally binding the protein of interest, can be administered to a mammal where they will associate with the half-life extender(*e.g.*, HSA, transferrin, IgG, red blood cells, *etc.*) to form a complex. This complex formation results in the half-life extension protecting the multimer and/or bound protein(s) from proteolytic degradation and/or other removal of the multimer and/or protein(s) and thereby extending the half-life of the protein and/or multimer (*see, e.g.*, example 3 below). One variation of this use of the invention includes the half-life extender-binding monomer covalently linked to the protein of interest. The protein of interest may include a monomer domain, a multimer of monomer domains, or a synthetic drug. Alternatively, monomers that bind to either immunoglobulins or erythrocytes could be generated using the above method and could be used for half-life extension.

[268] The half-life extender-binding multimers are typically multimers of at least two domains, chimeric domains, or mutagenized domains two domains, chimeric domains, or mutagenized domains (*i.e.*, one that binds to a target of interest and one that binds to the blood-carried molecule or cell). Suitable domains, *e.g.*, those described herein, can be further screened and selected for binding to a half-life extender. The half-life extender-binding multimers are generated in accordance with the methods for making multimers described herein, using, for example, monomer domains pre-screened for half-life extender -binding activity. For example, some half-life extender-binding LDL receptor class A-domain monomers are described in Example 2 below.

[269] In some embodiments, the multimers comprise at least one domain that binds to HSA, transferrin, IgG, a red blood cell or other half-life extender wherein the domain

comprises a trefoil/PD domain motif, a thrombospondin domain motif, or a thyroglobulin domain motif as provided herein, and the multimer comprises at least a second domain that binds a target molecule, wherein the second domain comprises a trefoil/PD domain motif, a thrombospondin domain motif, or a thyroglobulin domain motif as provided herein. The serum half-life of a molecule can be extended to be, *e.g.*, at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70 80, 90, 100, 150, 200, 250, 400, 500 or more hours.

[270] The present invention also provides a method for the suppression of or lowering of an immune response in a mammal. This method comprises first selecting a monomer domain that binds to an immunosuppressive target. Such an "immunosuppressive target" is defined as any protein that when bound by another protein produces an immunosuppressive result in a mammal. The immunosuppressive monomer domain can then be either administered directly or can be covalently linked to another monomer domain or to another protein that will provide the desired targeting of the immunosuppressive monomer. The immunosuppressive multimers are typically multimers of at least two domains, chimeric domains, or mutagenized domains. Suitable domains include all of those described herein and are further screened and selected for binding to an immunosuppressive target. Immunosuppressive multimers are generated in accordance with the methods for making multimers described herein, using, for example, trefoil/PD monomer domains, thrombospondin monomer domains, or thyroglobulin monomer domains.

[271] In another embodiment, a multimer comprising a first monomer domain that binds to the ligand and a second monomer domain that binds to the receptor can be used to increase the effective affinity of the ligand for the receptor.

[272] In another embodiment, multimers comprising at least two monomers that bind to receptors are used to bring two receptors into proximity by both binding the multimer, thereby activating the receptors.

[273] In some embodiments, multimers with two different monomers can be used to employ a target-driven avidity increase. For example, a first monomer can be targeted to a cell surface molecule on a first cell type and a second monomer can be targeted to a surface molecule on a second cell type. By linking the two monomers to forma a multimer and then adding the multimer to a mixture of the two cell types, binding will occur between the cells once an initial binding event occurs between one multimer and two cells, other multimers will also bind both cells.

[274] Further examples of potential uses of the invention include monomer domains, and multimers thereof, that are capable of drug binding (*e.g.*, binding

radionucleotides for targeting, pharmaceutical binding for half-life extension of drugs, controlled substance binding for overdose treatment and addiction therapy), immune function modulating (*e.g.*, immunogenicity blocking by binding such receptors as CTLA-4, immunogenicity enhancing by binding such receptors as CD80,or complement activation by Fc type binding), and specialized delivery (*e.g.*, slow release by linker cleavage, electrotransport domains, dimerization domains, or specific binding to: cell entry domains, clearance receptors such as FcR, oral delivery receptors such as pIgR for trans-mucosal transport, and blood-brain transfer receptors such as transferrinR).

[275] In further embodiments, monomers or multimers can be linked to a detectable label (*e.g.*, Cy3, Cy5, *etc.*) or linked to a reporter gene product (*e.g.*, CAT, luciferase, horseradish peroxidase, alkaline phosphotase, GFP, *etc.*).

[276] In some embodiments, the monomers of the invention are selected for the ability to bind antibodies from specific animals, *e.g.*, goat, rabbit, mouse, *etc.*, for use as a secondary reagent in detection assays.

[277] In some cases, a pair of monomers or multimers are selected to bind to the same target (*i.e.*, for use in sandwich-based assays). To select a matched monomer or multimer pair, two different monomers or multimers typically are able to bind the target protein simultaneously. One approach to identify such pairs involves the following:

(1) immobilizing the phage or protein mixture that was previously selected to bind the target protein

(2) contacting the target protein to the immobilized phage or protein and washing;

(3) contacting the phage or protein mixture to the bound target and washing; and

(4) eluting the bound phage or protein without eluting the immobilized phage or protein.

In some embodiments, different phage populations with different drug markers are used.

[278] One use of the multimers or monomer domains of the invention is use to replace antibodies or other affinity agents in detection or other affinity-based assays. Thus, in some embodiments, monomer domains or multimers are selected against the ability to bind components other than a target in a mixture. The general approach can include performing the affinity selection under conditions that closely resemble the conditions of the assay, including mimicking the composition of a sample during the assay. Thus, a step of selection could include contacting a monomer domain or multimer to a mixture not including the target ligand and selecting against any monomer domains or multimers that bind to the mixture. Thus, the mixtures (absent the target ligand, which could be depleted using an antibody,

monomer domain or multimer) representing the sample in an assay (serum, blood, tissue, cells, urine, semen, etc) can be used as a blocking agent. Such subtraction is useful, *e.g.*, to create pharmaceutical proteins that bind to their target but not to other serum proteins or non-target tissues.

X. Further Manipulating Monomer Domains and/or Multimer Nucleic Acids and Polypeptides

[279] As mentioned above, the polypeptide of the present invention can be altered. Descriptions of a variety of diversity generating procedures for generating modified or altered nucleic acid sequences encoding these polypeptides are described above and below in the following publications and the references cited therein: Soong *et al.*, (2000) Nat Genet 25(4):436-439; Stemmer, *et al.*, (1999) Tumor Targeting 4:1-4; Ness *et al.*, (1999) Nat. Biotech. 17:893-896; Chang *et al.*, (1999) Nat. Biotech. 17:793-797; Minshull and Stemmer, (1999) Curr. Op. Chem. Biol. 3:284-290; Christians *et al.*, (1999) Nat. Biotech. 17:259-264; Crameri *et al.*, (1998) Nature 391:288-291; Crameri *et al.*, (1997) Nat. Biotech. 15:436-438; Zhang *et al.*, (1997) PNAS USA 94:4504-4509; Patten *et al.*, (1997) Curr. Op. Biotech. 8:724-733; Crameri *et al.*, (1996) Nat. Med. 2:100-103; Crameri *et al.*, (1996) Nat. Biotech. 14:315-319; Gates *et al.*, (1996) J. Mol. Biol. 255:373-386; Stemmer, (1996) In: The Encyclopedia of Molecular Biology. VCH Publishers, New York. pp.447-457; Crameri and Stemmer, (1995) BioTechniques 18:194-195; Stemmer *et al.*, (1995) Gene, 164:49-53; Stemmer, (1995) Science 270: 1510; Stemmer, (1995) Bio/Technology 13:549-553; Stemmer, (1994) Nature 370:389-391; and Stemmer, (1994) PNAS USA 91:10747-10751.

[280] Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling *et al.*, (1997) Anal Biochem. 254(2): 157-178; Dale *et al.*, (1996) Methods Mol. Biol. 57:369-374; Smith, (1985) Ann. Rev. Genet. 19:423-462; Botstein & Shortle, (1985) Science 229:1193-1201; Carter, (1986) Biochem. J. 237:1-7; and Kunkel, (1987) in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel, (1985) PNAS USA 82:488-492; Kunkel *et al.*, (1987) Methods in Enzymol. 154, 367-382; and Bass *et al.*, (1988) Science 242:240-245); oligonucleotide-directed mutagenesis ((1983) Methods in Enzymol. 100: 468-500; (1987) Methods in Enzymol. 154: 329-350; Zoller & Smith, (1982) Nucleic Acids Res. 10:6487-6500; Zoller & Smith, (1983) Methods in Enzymol. 100:468-500; and Zoller & Smith, (1987) Methods in Enzymol. 154:329-350); phosphorothioate-modified

DNA mutagenesis (Taylor *et al.*, (1985) Nucl. Acids Res. 13: 8749-8764; Taylor *et al.*, (1985) Nucl. Acids Res. 13: 8765-8787; Nakamaye & Eckstein, (1986) Nucl. Acids Res. 14: 9679-9698; Sayers *et al.*, (1988) Nucl. Acids Res. 16:791-802; and Sayers *et al.*, (1988) Nucl. Acids Res. 16: 803-814); mutagenesis using gapped duplex DNA (Kramer *et al.*, (1984) Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. 154:350-367; Kramer *et al.*, (1988) Nucl. Acids Res. 16: 7207; and Fritz *et al.*, (1988) Nucl. Acids Res. 16: 6987-6999).

[281] Additional suitable methods include point mismatch repair (Kramer *et al.*, *Point Mismatch Repair*, (1984) Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter *et al.*, (1985) Nucl. Acids Res. 13: 4431-4443; and Carter, (1987) Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh & Henikoff, (1986) Nucl. Acids Res. 14: 5115), restriction-selection and restriction-purification (Wells *et al.*, (1986) Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar *et al.*, (1984) Science 223: 1299-1301; Sakamar and Khorana, (1988) Nucl. Acids Res. 14: 6361-6372; Wells *et al.*, (1985) Gene 34:315-323; and Grundström *et al.*, (1985) Nucl. Acids Res. 13: 3305-3316), double-strand break repair (Mandecki, (1986) PNAS USA, 83:7177-7181; and Arnold, (1993) Curr. Op. Biotech. 4:450-455). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[282] Additional details regarding various diversity generating methods can be found in U.S. Patent Nos. 5,605,793; 5,811,238; 5,830,721; 5,834,252; 5,837,458; WO 95/22625; WO 96/33207; WO 97/20078; WO 97/35966; WO 99/41402; WO 99/41383; WO 99/41369; WO 99/41368; EP 752008; EP 0932670; WO 99/23107; WO 99/21979; WO 98/31837; WO 98/27230; WO 98/27230; WO 00/00632; WO 00/09679; WO 98/42832; WO 99/29902; WO 98/41653; WO 98/41622; WO 98/42727; WO 00/18906; WO 00/04190; WO 00/42561; WO 00/42559; WO 00/42560; WO 01/23401; PCT/US01/06775.

[283] Another aspect of the present invention includes the cloning and expression of monomer domains, selected monomer domains, multimers and/or selected multimers coding nucleic acids. Thus, multimer domains can be synthesized as a single protein using expression systems well known in the art. In addition to the many texts noted above, general texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other topics relevant to expressing nucleic acids such as monomer domains, selected monomer domains, multimers and/or selected multimers, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in

Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel")). Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, useful in identifying, isolating and cloning monomer domains and multimers coding nucleic acids, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q-replicase amplification and other RNA polymerase mediated techniques (*e.g.*, NASBA), are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3, 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86, 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem* 35, 1826; Landegren *et al.*, (1988) *Science* 241, 1077-1080; Van Brunt (1990) *Biotechnology* 8, 291-294; Wu and Wallace, (1989) *Gene* 4, 560; Barringer *et al.* (1990) *Gene* 89, 117, and Sooknanan and Malek (1995) *Biotechnology* 13: 563-564. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in Cheng *et al.* (1994) *Nature* 369: 684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See*, Ausubel, Sambrook and Berger, *all supra.*

[284] The present invention also relates to the introduction of vectors of the invention into host cells, and the production of monomer domains, selected monomer domains immuno-domains, multimers and/or selected multimers of the invention by recombinant techniques. Host cells are genetically engineered (*i.e.*, transduced, transformed or transfected) with the vectors of this invention, which can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, a viral particle, a phage, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the monomer domain, selected monomer domain, multimer and/or selected multimer gene(s) of interest. The culture conditions, such as temperature, pH and the like,

are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, *e.g.*, Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, third edition, Wiley- Liss, New York and the references cited therein.

[285] As mentioned above, the polypeptides of the invention can also be produced in non-animal cells such as plants, yeast, fungi, bacteria and the like. Indeed, as noted throughout, phage display is an especially relevant technique for producing such polypeptides. In addition to Sambrook, Berger and Ausubel, details regarding cell culture can be found in Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture*; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

[286] The present invention also includes alterations of monomer domains, immuno-domains and/or multimers to improve pharmacological properties, to reduce immunogenicity, or to facilitate the transport of the multimer and/or monomer domain into a cell or tissue (*e.g.*, through the blood-brain barrier, or through the skin). These types of alterations include a variety of modifications (*e.g.*, the addition of sugar-groups or glycosylation), the addition of PEG, the addition of protein domains that bind a certain protein (*e.g.*, HSA or other serum protein), the addition of proteins fragments or sequences that signal movement or transport into, out of and through a cell. Additional components can also be added to a multimer and/or monomer domain to manipulate the properties of the multimer and/or monomer domain. A variety of components can also be added including, *e.g.*, a domain that binds a known receptor (*e.g.*, a Fc-region protein domain that binds a Fc receptor), a toxin(s) or part of a toxin, a prodomain that can be optionally cleaved off to activate the multimer or monomer domain, a reporter molecule (*e.g.*, green fluorescent protein), a component that bind a reporter molecule (such as a radionuclide for radiotherapy, biotin or avidin) or a combination of modifications.

XI. Additional Methods of Screening

[287] The present invention also provides a method for screening a protein for potential immunogenicity by:

providing a candidate protein sequence;

comparing the candidate protein sequence to a database of human protein sequences;

identifying portions of the candidate protein sequence that correspond to portions of human protein sequences from the database; and

determining the extent of correspondence between the candidate protein sequence and the human protein sequences from the database.

[288] In general, the greater the extent of correspondence between the candidate protein sequence and one or more of the human protein sequences from the database, the lower the potential for immunogenicity is predicted as compared to a candidate protein having little correspondence with any of the human protein sequences from the database. Removal or limitation of the number of immunogenic amino acids and/or sequences may also be used to reduce immunogenicity of the monomer domains, *e.g.*, either before or after the libraries are screened. Immunogenic sequences include, *e.g.*, HLA type I or type II sequences or proteasome sites. A variety of commercial products and computer programs are available to identify these amino acids, *e.g.*, Tepitope (Roche), the Parker Matrix, ProPred-I matrix, Biovation, Epivax, Epimatrix.

[289] A database of human protein sequences that is suitable for use in the practice of the invention method for screening candidate proteins can be found at ncbi.nlm.nih.gov/blast/Blast.cgi at the World Wide Web (in addition, the following web site can be used to search short, nearly exact matches: cbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=(none)&EXPECT=1000&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=29&SET_DEFAULTS.y=6&SHOW_OVERVIEW=on&WORD_SIZE=7&END_OF_HTTPGET=Yes&SHOW_LINKOUT=yes at the World Wide Web). The method is particularly useful in determining whether a crossover sequence in a chimeric protein, such as, for example, a chimeric monomer domain, is likely to cause an immunogenic event. If the crossover sequence corresponds to a portion of a sequence found in the database of human protein sequences, it is believed that the crossover sequence is less likely to cause an immunogenic event.

[290] Human chimeric domain libraries prepared in accordance to the methods of the present invention can be screened for potential immunogenicity, in addition to binding affinity. Furthermore, information pertaining to portions of human protein sequences from the database can be used to design a protein library of human-like chimeric proteins.

Such library can be generated by using information pertaining to "crossover sequences" that exist in naturally occurring human proteins. The term "crossover sequence" refers herein to a sequence that is found in its entirety in at least one naturally occurring human protein, in which portions of the sequence are found in two or more naturally occurring proteins. Thus, recombination of the latter two or more naturally occurring proteins would generate a chimeric protein in which the chimeric portion of the sequence actually corresponds to a sequence found in another naturally occurring protein. The crossover sequence contains a chimeric junction of two consecutive amino acid residue positions in which the first amino acid position is occupied by an amino acid residue identical in type and position found in a first and second naturally occurring human protein sequence, but not a third naturally occurring human protein sequence. The second amino acid position is occupied by an amino acid residue identical in type and position found in a second and third naturally occurring human protein sequence, but not the first naturally occurring human protein sequence. In other words, the "second" naturally occurring human protein sequence corresponds to the naturally occurring human protein in which the crossover sequence appears in its entirety, as described above.

[291] In accordance with the present invention, a library of human-like chimeric proteins is generated by: identifying human protein sequences from a database that correspond to proteins from the same family of proteins; aligning the human protein sequences from the same family of proteins to a reference protein sequence; identifying a set of subsequences derived from different human protein sequences of the same family, wherein each subsequence shares a region of identity with at least one other subsequence derived from a different naturally occurring human protein sequence; identifying a chimeric junction from a first, a second, and a third subsequence, wherein each subsequence is derived from a different naturally occurring human protein sequence, and wherein the chimeric junction comprises two consecutive amino acid residue positions in which the first amino acid position is occupied by an amino acid residue common to the first and second naturally occurring human protein sequence, but not the third naturally occurring human protein sequence, and the second amino acid position is occupied by an amino acid residue common to the second and third naturally occurring human protein sequence, and generating human-like chimeric protein molecules each corresponding in sequence to two or more subsequences from the set of subsequences, and each comprising one of more of the identified chimeric junctions.

[292] Thus, for example, if the first naturally occurring human protein sequence is, A-B-C, and the second is, B-C-D-E, and the third is, D-E-F, then the chimeric junction is C-D. Alternatively, if the first naturally occurring human protein sequence is D-E-F-G, and the second is B-C-D-E-F, and the third is A-B-C-D, then the chimeric junction is D-E. Human-like chimeric protein molecules can be generated in a variety of ways. For example, oligonucleotides comprising sequences encoding the chimeric junctions can be recombined with oligonucleotides corresponding in sequence to two or more subsequences from the above-described set of subsequences to generate a human-like chimeric protein, and libraries thereof. The reference sequence used to align the naturally occurring human proteins is a sequence from the same family of naturally occurring human proteins, or a chimera or other variant of proteins in the family.

XII. Animal Models

[293] Another aspect of the invention is the development of specific non-human animal models in which to test the immunogenicity of the monomer or multimer domains. The method of producing such non-human animal model comprises: introducing into at least some cells of a recipient non-human animal, vectors comprising genes encoding a plurality of human proteins from the same family of proteins, wherein the genes are each operably linked to a promoter that is functional in at least some of the cells into which the vectors are introduced such that a genetically modified non-human animal is obtained that can express the plurality of human proteins from the same family of proteins.

[294] Suitable non-human animals employed in the practice of the present invention include all vertebrate animals, except humans (*e.g.*, mouse, rat, rabbit, sheep, and the like). Typically, the plurality of members of a family of proteins includes at least two members of that family, and usually at least ten family members. In some embodiments, the plurality includes all known members of the family of proteins. Exemplary genes that can be used include those encoding monomer domains, such as, for example, members of the thrombospondin type I domain family, thyroglobulin domain family, or trefoil domain family, as well as the other domain families described herein.

[295] The non-human animal models of the present invention can be used to screen for immunogenicity of a monomer or multimer domain that is derived from the same family of proteins expressed by the non-human animal model. The present invention includes the non-human animal model made in accordance with the method described above,

as well as transgenic non-human animals whose somatic and germ cells contain and express DNA molecules encoding a plurality of human proteins from the same family of proteins (such as the monomer domains described herein), wherein the DNA molecules have been introduced into the transgenic non-human animal at an embryonic stage, and wherein the DNA molecules are each operably linked to a promoter in at least some of the cells in which the DNA molecules have been introduced.

[296] An example of a mouse model useful for screening thrombospondin type I domain, thyroglobulin domain, or trefoil domain derived binding proteins is described as follows. Gene clusters encoding the wild type human thrombospondin type I monomer domains, thyroglobulin monomer domains, or trefoil monomer domains are amplified from human cells using PCR. These fragments are then used to generate transgenic mice according to the method described above. The transgenic mice will recognize the human thrombospondin type I domains, thyroglobulin domains, or trefoil domains as "self", thus mimicking the "selfness" of a human with regard to thrombospondin type I domains, thyroglobulin domains, or trefoil domains. Individual thrombospondin type I derived monomers, thyroglobulin derived monomers, or trefoil derived monomers or multimers are tested in these mice by injecting the thrombospondin type I derived monomers or multimers, thyroglobulin derived monomers or multimers, or trefoil derived monomers or multimers into the mice, then analyzing the immune response (or lack of response) generated. The mice are tested to determine if they have developed a mouse anti-human response (MAHR). Monomers and multimers that do not result in the generation of a MAHR are likely to be non-immunogenic when administered to humans.

[297] Historically, MAHR test in transgenic mice is used to test individual proteins in mice that are transgenic for that single protein. In contrast, the above described method provides a non-human animal model that recognizes an entire family of human proteins as "self," and that can be used to evaluate a huge number of variant proteins that each are capable of vastly varied binding activities and uses.

XIII. Kits

[298] Kits comprising the components needed in the methods (typically in an unmixed form) and kit components (packaging materials, instructions for using the components and/or the methods, one or more containers (reaction tubes, columns, *etc.*)) for holding the components are a feature of the present invention. Kits of the present invention

may contain a multimer library, or a single type of multimer. Kits can also include reagents suitable for promoting target molecule binding, such as buffers or reagents that facilitate detection, including detectably-labeled molecules. Standards for calibrating a ligand binding to a monomer domain or the like, can also be included in the kits of the invention.

[299] The present invention also provides commercially valuable binding assays and kits to practice the assays. In some of the assays of the invention, one or more ligand is employed to detect binding of a monomer domain, immuno-domains and/or multimer. Such assays are based on any known method in the art, *e.g.*, flow cytometry, fluorescent microscopy, plasmon resonance, and the like, to detect binding of a ligand(s) to the monomer domain and/or multimer.

[300] Kits based on the assay are also provided. The kits typically include a container, and one or more ligand. The kits optionally comprise directions for performing the assays, additional detection reagents, buffers, or instructions for the use of any of these components, or the like. Alternatively, kits can include cells, vectors, (*e.g.*, expression vectors, secretion vectors comprising a polypeptide of the invention), for the expression of a monomer domain and/or a multimer of the invention.

[301] In a further aspect, the present invention provides for the use of any composition, monomer domain, immuno-domain, multimer, cell, cell culture, apparatus, apparatus component or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein and/or for the use of cells, cell cultures, compositions or other features herein as a therapeutic formulation. The manufacture of all components herein as therapeutic formulations for the treatments described herein is also provided.

XIV. Integrated Systems

[302] The present invention provides computers, computer readable media and integrated systems comprising character strings corresponding to monomer domains, selected monomer domains, multimers and/or selected multimers and nucleic acids encoding such polypeptides. These sequences can be manipulated by in silico recombination methods, or by standard sequence alignment or word processing software.

[303] For example, different types of similarity and considerations of various stringency and character string length can be detected and recognized in the integrated systems herein. For example, many homology determination methods have been designed

for comparative analysis of sequences of biopolymers, for spell checking in word processing, and for data retrieval from various databases. With an understanding of double-helix pair-wise complement interactions among 4 principal nucleobases in natural polynucleotides, models that simulate annealing of complementary homologous polynucleotide strings can also be used as a foundation of sequence alignment or other operations typically performed on the character strings corresponding to the sequences herein (*e.g.*, word-processing manipulations, construction of figures comprising sequence or subsequence character strings, output tables, *etc.*). An example of a software package with GOs for calculating sequence similarity is BLAST, which can be adapted to the present invention by inputting character strings corresponding to the sequences herein.

[304] BLAST is described in Altschul *et al.*, (1990) J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (available on the World Wide Web at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[305] An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, (1987) J. Mol. Evol. 35:351-360. The method used is similar to the method described by Higgins & Sharp, (1989) CABIOS 5:151-153. The program can align, *e.g.*, up to 300 sequences of a maximum length of 5,000 letters. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program can also be used to plot a dendogram or tree representation of clustering relationships. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison. For example, in order to determine conserved amino acids in a monomer domain family or to compare the sequences of monomer domains in a family, the sequence of the invention, or coding nucleic acids, are aligned to provide structure-function information.

[306] In one aspect, the computer system is used to perform "in silico" sequence recombination or shuffling of character strings corresponding to the monomer domains. A variety of such methods are set forth in "Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics" by Selifonov and Stemmer, filed February 5, 1999 (USSN 60/118854) and "Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics" by Selifonov and Stemmer, filed October 12, 1999 (USSN 09/416,375). In brief, genetic operators are used in genetic algorithms to change given sequences, *e.g.*, by mimicking genetic events such as mutation, recombination, death and the like. Multi-dimensional analysis to optimize sequences can be also be performed in the computer system, *e.g.*, as described in the '375 application.

[307] A digital system can also instruct an oligonucleotide synthesizer to synthesize oligonucleotides, *e.g.*, used for gene reconstruction or recombination, or to order oligonucleotides from commercial sources (*e.g.*, by printing appropriate order forms or by linking to an order form on the Internet).

[308] The digital system can also include output elements for controlling nucleic acid synthesis (*e.g.*, based upon a sequence or an alignment of a recombinant, *e.g.*,

recombined, monomer domain as herein), *i.e.*, an integrated system of the invention optionally includes an oligonucleotide synthesizer or an oligonucleotide synthesis controller. The system can include other operations that occur downstream from an alignment or other operation performed using a character string corresponding to a sequence herein, *e.g.*, as noted above with reference to assays.

EXAMPLES

[309] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1

[310] This example describes selection of monomer domains and the creation of multimers.

[311] Starting materials for identifying monomer domains and creating multimers from the selected monomer domains and procedures can be derived from any of a variety of human and/or non-human sequences. For example, to produce a selected monomer domain with specific binding for a desired ligand or mixture of ligands, one or more monomer domain gene(s) are selected from a family of monomer domains that bind to a certain ligand. The nucleic acid sequences encoding the one or more monomer domain gene can be obtained by PCR amplification of genomic DNA or cDNA, or optionally, can be produced synthetically using overlapping oligonucleotides.

[312] Most commonly, these sequences are then cloned into a cell surface display format (*i.e.*, bacterial, yeast, or mammalian (COS) cell surface display; phage display) for expression and screening. The recombinant sequences are transfected (transduced or transformed) into the appropriate host cell where they are expressed and displayed on the cell surface. For example, the cells can be stained with a labeled (*e.g.*, fluorescently labeled), desired ligand. The stained cells are sorted by flow cytometry, and the selected monomer domains encoding genes are recovered (*e.g.*, by plasmid isolation, PCR or expansion and cloning) from the positive cells. The process of staining and sorting can be repeated multiple times (*e.g.*, using progressively decreasing concentrations of the desired ligand until a desired level of enrichment is obtained). Alternatively, any screening or detection method known in the art that can be used to identify cells that bind the desired ligand or mixture of ligands can be employed.

[313] The selected monomer domain encoding genes recovered from the desired ligand or mixture of ligands binding cells can be optionally recombined according to

any of the methods described herein or in the cited references. The recombinant sequences produced in this round of diversification are then screened by the same or a different method to identify recombinant genes with improved affinity for the desired or target ligand. The diversification and selection process is optionally repeated until a desired affinity is obtained.

[314] The selected monomer domain nucleic acids selected by the methods can be joined together via a linker sequence to create multimers, *e.g.*, by the combinatorial assembly of nucleic acid sequences encoding selected monomer domains by DNA ligation, or optionally, PCR-based, self-priming overlap reactions. The nucleic acid sequences encoding the multimers are then cloned into a cell surface display format (*i.e.*, bacterial, yeast, or mammalian (COS) cell surface display; phage display) for expression and screening. The recombinant sequences are transfected (transduced or transformed) into the appropriate host cell where they are expressed and displayed on the cell surface. For example, the cells can be stained with a labeled, *e.g.*, fluorescently labeled, desired ligand or mixture of ligands. The stained cells are sorted by flow cytometry, and the selected multimers encoding genes are recovered (*e.g.*, by PCR or expansion and cloning) from the positive cells. Positive cells include multimers with an improved avidity or affinity or altered specificity to the desired ligand or mixture of ligands compared to the selected monomer domain(s). The process of staining and sorting can be repeated multiple times (*e.g.*, using progressively decreasing concentrations of the desired ligand or mixture of ligands until a desired level of enrichment is obtained). Alternatively, any screening or detection method known in the art that can be used to identify cells that bind the desired ligand or mixture of ligands can be employed.

[315] The selected multimer encoding genes recovered from the desired ligand or mixture of ligands binding cells can be optionally recombined according to any of the methods described herein or in the cited references. The recombinant sequences produced in this round of diversification are then screened by the same or a different method to identify recombinant genes with improved avidity or affinity or altered specificity for the desired or target ligand. The diversification and selection process is optionally repeated until a desired avidity or affinity or altered specificity is obtained.

Example 2

[316] This example describes the selection of monomer domains that are capable of binding to Human Serum Albumin (HSA).

[317] For the production of phages, *E. coli* DH10B cells (Invitrogen) were transformed with phage vectors encoding a library of LDL receptor class A-domain variants

as a fusions to the pIII phage protein. To transform these cells, the electroporation system MicroPulser (Bio-Rad) was used together with cuvettes provided by the same manufacturer. The DNA solution was mixed with 100 µl of the cell suspension, incubated on ice and transferred into the cuvette (electrode gap 1mm). After pulsing, 2 ml of SOC medium (2 % w/v tryptone, 0.5 % w/v yeast extract, 10 mM NaCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$) were added and the transformation mixture was incubated at 37 C for 1 h. Multiple transformations were combined and diluted in 500 ml 2xYT medium containing 20 µg/m tetracycline and 2 mM $CaCl_2$. With 10 electroporations using a total of 10 µg ligated DNA $1.2x10^8$ independent clones were obtained.

[318] 160 ml of the culture, containing the cells which were transformed with the phage vectors encoding the library of the A-domain variant phages, were grown for 24 h at 22 C, 250 rpm and afterwards transferred in sterile centrifuge tubes. The cells were sedimented by centrifugation (15 minutes, 5000 g, 4 °C). The supernatant containing the phage particles was mixed with 1/5 volumes 20 % w/v PEG 8000, 15 % w/v NaCl, and was incubated for several hours at 4 °C. After centrifugation (20 minutes, 10000 g, 4 °C) the precipitated phage particles were dissolved in 2 ml of cold TBS (50 mM Tris, 100 mM NaCl, pH 8.0) containing 2 mM $CaCl_2$. The solution was incubated on ice for 30 minutes and was distributed into two 1.5 ml reaction vessels. After centrifugation to remove undissolved components (5 minutes, 18500 g, 4 °C) the supernatants were transferred to a new reaction vessel. Phage were reprecipitated by adding 1/5 volumes 20 % w/v PEG 8000, 15 % w/v NaCl and incubation for 60 minutes on ice. After centrifugation (30 minutes, 18500 g, 4 °C) and removal of the supernatants, the precipitated phage particles were dissolved in a total of 1 ml TBS containing 2 mM $CaCl_2$. After incubation for 30 minutes on ice the solution was centrifuged as described above. The supernatant containing the phage particles was used directly for the affinity enrichment.

[319] Affinity enrichment of phage was performed using 96 well plates (Maxisorp, NUNC, Denmark). Single wells were coated for 12 h at RT by incubation with 150 µl of a solution of 100 µg/ml human serum albumin (HSA, Sigma) in TBS. Binding sites remaining after HSA incubation were saturated by incubation with 250 µl 2% w/v bovine serum albumin (BSA) in TBST (TBS with 0.1 % v/v Tween 20) for 2 hours at RT. Afterwards, 40 µl of the phage solution, containing approximately $5x10^{11}$ phage particles, were mixed with 80 µl TBST containing 3 % BSA and 2 mM $CaCl_2$ for 1 hour at RT. In

order to remove non binding phage particles, the wells were washed 5 times for 1 min using 130 µl TBST containing 2 mM $CaCl_2$.

[320] Phage bound to the well surface were eluted either by incubation for 15 minutes with 130 µl 0.1 M glycine/HCl pH 2.2 or in a competitive manner by adding 130 µl of 500 µg/ml HSA in TBS. In the first case, the pH of the elution fraction was immediately neutralized after removal from the well by mixing the eluate with 30 µl 1 M Tris/HCl pH 8.0.

[321] For the amplification of phage, the eluate was used to infect *E. coli* K91BluKan cells (F^+). 50 µl of the eluted phage solution were mixed with 50 µl of a preparation of cells and incubated for 10 minutes at RT. Afterwards, 20 ml LB medium containing 20 µg/ml tetracycline were added and the infected cells were grown for 36 h at 22 C, 250 rpm. Afterwards, the cells were sedimented (10 minutes, 5000 g, 4 °C). Phage were recovered from the supernatant by precipitation as described above. For the repeated affinity enrichment of phage particles the same procedure as described in this example was used. After two subsequent rounds of panning against HSA, random colonies were picked and tested for their binding properties against the used target protein.

[322] While this example demonstrates the use of LDL-receptor A domains, those of skill in the art will appreciate that the same techniques can be used to generate desired binding properties in monomer domains of the present invention.

Example 3

[323] This example describes the determination of biological activity of monomer domains that are capable of binding to HSA.

[324] In order to show the ability of an HSA binding domain to extend the serum half life of an protein *in vivo*, the following experimental setup was performed. A multimeric A-domain, consisting of an A-domain which was evolved for binding HSA (see Example 2) and a streptavidin binding A-domain was compared to the streptavidin binding A-domain itself. The proteins were injected into mice, which were either loaded or not loaded (as control) with human serum albumin (HSA). Serum levels of a-domain proteins were monitored.

[325] Therefore, an A-domain, which was evolved for binding HSA (see Example 1) was fused on the genetic level with a streptavidin binding A-domain multimer using standard molecular biology methods (see Maniatis *et al.*). The resulting genetic construct, coding for an A-domain multimer as well as a hexahistidine tag and a HA tag, were

used to produce protein in *E. coli*. After refolding and affinity tag mediated purification the proteins were dialysed several times against 150 mM NaCl, 5 mM Tris pH 8.0, 100 μM $CaCl_2$ and sterile filtered (0.45 μM).

[326] Two sets of animal experiments were performed. In a first set, 1 ml of each prepared protein solution with a concentration of 2.5 μM were injected into the tail vein of separate mice and serum samples were taken 2, 5 and 10 minutes after injection. In a second set, the protein solution described before was supplemented with 50 mg/ml human serum albumin. As described above, 1 ml of each solution was injected per animal. In case of the injected streptavidin binding A-domain dimer, serum samples were taken 2, 5 and 10 minutes after injection, while in case of the trimer, serum samples were taken after 10, 30 and 120 minutes. All experiments were performed as duplicates and individual animals were assayed per time point.

[327] In order to detect serum levels of A-domains in the serum samples, an enzyme linked immunosorbent assay (ELISA) was performed. Therefore, wells of a maxisorp 96 well microtiter plate (NUNC, Denmark) were coated with each 1 μg anti-His_6-antibody in TBS containing 2 mM $CaCl_2$ for 1 h at 4 C. After blocking remaining binding sites with casein (Sigma) solution for 1 h, wells were washed three times with TBS containing 0.1 % Tween and 2 mM $CaCl_2$. Serial concentration dilutions of the serum samples were prepared and incubated in the wells for 2 h in order to capture the a-domain proteins. After washing as before, anti-HA-tag antibody coupled to horse radish peroxidase (HRP) (Roche Diagnostics, 25 μg/ml) was added and incubated for 2 h. After washing as described above, HRP substrate (Pierce) was added and the detection reaction developed according to the instructions of the manufacturer. Light absorption, reflecting the amount of a-domain protein present in the serum samples, was measured at a wavelength of 450 nm. Obtained values were normalized and plotted against a time scale.

[328] Evaluation of the obtained values showed a serum half life for the streptavidin binding A-domain of about 4 minutes without presence of HSA respectively 5.2 minutes when the animal was loaded with HSA. The trimer of A-domains, which contained the HSA binding A-domain, exhibited a serum half life of 6.3 minutes without the presence of HSA but a significantly increased half life of 38 minutes when HSA was present in the animal. This clearly indicates that the HSA binding A-domain can be used as a fusion partner to increase the serum half life of any protein, including protein therapeuticals.

Example 4

[329] This example describes experiments demonstrating extension of half-life of proteins in blood.

[330] To further demonstrate that blood half-life of proteins can be extended using monomer domains of the invention, individual monomer domain proteins selected against monkey serum albumin, human serum albumin, human IgG, and human red blood cells were added to aliquots of whole, heparinized human or monkey blood.

[331] The following list provides sequences of monomer domains analyzed in this example.

```
IG156   CLSSEFQCQSSGRCIPLAWVCDGDNDCRDDSDEKSCKPRT
RBCA    CRSSQFQCNDSRICIPGRWRCDGDNDCQDGSDETGCGDSHILPFSTPGPST
RBCB    CPAGEFPCKNGQCLPVTWLCDGVNDCLDGSDEKGCGRPGPGATSAPAA
RBC11   CPPDEFPCKNGQCIPQDWLCDGVNDCLDGSDEKDCGRPGPGATSAPAA
CSA-A8  CGAGQFPCKNGHCLPLNLLCDGVNDCEDNSDEPSELCKALT
```

[332] Blood aliquots containing monomer protein were then added to individual dialysis bags (25,000 MWCO), sealed, and stirred in 4 L of Tris-buffered saline at room temperature overnight.

[333] Anti-6xHis antibody was immobilized by hydrophobic interaction to a 96-well plate (Nunc). Serial dilutions of serum from each blood sample were incubated with the immobilized antibody for 3 hours. Plates were washed to remove unbound protein and probed with α-HA-HRP to detect monomer.

[334] Monomers identified as having long half-lives in dialysis experiments were constructed to contain either an HA, FLAG, E-Tag, or myc epitope tag. Four monomers were pooled, containing one protein for each tag, to make two pools.

[335] One monkey was injected subcutaneously per pool, at a dose of 0.25 mg/kg/monomer in 2.5 mL total volume in saline. Blood samples were drawn at 24, 48, 96, and 120 hours. Anti-6xHis antibody was immobilized by hydrophobic interaction to a 96-well plate (Nunc). Serial dilutions of serum from each blood sample were incubated with the immobilized antibody for 3 hours. Plates were washed to remove unbound protein and separately probed with α-HA-HRP, α-FLAG-HRP, α-ETag-HRP, andα-myc-HRP to detect the monomer.

[336] The following illustrates a comparison between commercial antibodies and an anti-IgG multimer:

Drug		Mol. Wt.	Human T1/2	Dosing
Rebif	rIFN-b	23 kD	69 hrs	Weekly 3x
Pegasys	rIFN-a-PEG	40 kD	78 hrs	Weekly
Rituxan	CD20 Antibody	150 kD	78 hrs	Weekly
Enbrel	sTNF-R-Fc	150 kD	103 hrs	Weekly 2x
Multimer	Anti-IgG	5 kD	120 hrs	Weekly 1-2x
Herceptin	Her2 Antibody	150 kD	144 hrs	Weekly
Remicade	TNFa Antibody	150 kD	216 hrs	Monthly .5x
Humira	TNFa Antibody	150 kD	336 hrs	Monthly 2x

Example 5

[337] This example describes the development of protein-specific monomer domains and dimers by "walking."

[338] A library of DNA sequences encoding monomeric domains is created by assembly PCR as described in Stemmer *et al.*, *Gene* 164:49-53 (1995).

[339] PCR fragments were digested with appropriate restriction enzymes (*e.g.*, XmaI and SfiI). Digestion products were separated on 3% agarose gel and domain fragments are purified from the gel. The DNA fragments are ligated into the corresponding restriction sites of phage display vector fuse5-HA, a derivative of fuse5 carrying an in-frame HA-epitope. The ligation mixture is electroporated into TransforMax™ EC100™ electrocompetent *E. coli* cells. Transformed *E. coli* cells are grown overnight at 37°C in 2xYT medium containing 20 μg/ml tetracycline and 2 mM $CaCl_2$.

[340] Phage particles are purified from the culture medium by PEG-precipitation. Individual wells of a 96-well microtiter plate (Maxisorp) are coated with target protein (1 μg/well) in 0.1 M $NaHCO_3$. After blocking the wells with TBS buffer containing 10 mg/ml casein, purified phage is added at a typical number of ~$1–3 \times 10^{11}$. The microtiter plate is incubated at 4°C for 4 hours, washed 5 times with washing buffer (TBS/Tween) and bound phages are eluted by adding glycine-HCl buffer pH 2.2. The eluate is neutralized by adding 1 M Tris-HCl (pH 9.1). The phage eluate is amplified using *E. coli* K91BlueKan cells

and after purification used as input to a second and a third round of affinity selection (repeating the steps above).

[341] Phage from the final eluate is used directly, without purification, as a template to PCR amplify domain encoding DNA sequences.

[342] The PCR products are purified and subsequently digested with suitable restriction enzymes (*e.g.*, 50% with BpmI and 50% with BsrDI).

[343] The digested monomer fragments are 'walked' to dimers by attaching a library of naive domain fragments using DNA ligation. Naive domain sequences are obtained by PCR amplification of the initial domain library (resulting from the PEG purification described above) using primers suitable for amplifying the domains. The PCR fragments are purified, split into 2 equal amounts and then digested with suitable restriction enzymes (*e.g.*, either BpmI or BsrDI).

[344] Digestion products are separated on a 2% agarose gel and domain fragments were purified from the gel. The purified fragments are combined into 2 separate pools (*e.g.*, naïve/BpmI + selected/BsrDI & naïve/BsrDI + selected/BpmI) and then ligated overnight at 16°C.

[345] The dimeric domain fragments are PCR amplified (5 cycles), digested with suitable restriction enzymes (*e.g.*, XmaI and SfiI) and purified from a 2% agarose gel. Screening steps are repeated as described above except for the washing, which is done more stringently to obtain high-affinity binders. After infection, the K91BlueKan cells are plated on 2xYT agar plates containing 40 μg/ml tetracycline and grown overnight. Single colonies are picked and grown overnight in 2xYT medium containing 20 μg/ml tetracycline and 2 mM $CaCl_2$. Phage particles are purified from these cultures.

[346] Binding of the individual phage clones to their target proteins was analyzed by ELISA. Clones yielding the highest ELISA signals were sequenced and subsequently recloned into a protein expression vector.

[347] Protein production is induced in the expression vectors with IPTG and purified by metal chelate affinity chromatography. Protein-specific monomers are characterized as follows.

Biacore

[348] Two hundred fifty RU protein are immobilized by NHS/EDC coupling to a CM5 chip (Biacore). 0.5 and 5 μM solutions of monomer protein are flowed over the derivatized chip, and the data is analyzed using the standard Biacore software package.

ELISA

[349] Ten nanograms of protein per well is immobilized by hydrophobic interaction to 96-well plates (Nunc). Plates were blocked with 5 mg/mL casein. Serial dilutions of monomer protein were added to each well and incubated for 3 hours. Plates were washed to remove unbound protein and probed with α-HA-HRP to detect monomers.

Functional Assays

[350] Functional assays to determine the biological activity of the monomers can also be conducted and include, *e.g.*, assays to determine the binding specificity of the monomers, assays to determine whether the monomers antagonize or stimulate a metabolic pathway by binding to their target molecule, and the like.

Example 6

[351] This example describes *in vivo* intra-protein recombination to generate libraries of greater diversity.

[352] A monomer-encoding plasmid vector (pCK-derived vector; see below), flanked by orthologous *loxP* sites, was recombined in a Cre-dependent manner with a phage vector via its compatible *loxP* sites. The recombinant phage vectors were detected by PCR using primers specific for the recombinant construct. DNA sequencing indicated that the correct recombinant product was generated.

Reagents and experimental procedures

[353] **pCK-*cre*-lox-Mb-loxP.** This vector has two particularly relevant features. First, it carries the *cre* gene, encoding the site-specific DNA recombinase Cre, under the control of P_{lac}. *Cre* was PCR-amplified from p705-*cre* (from *GeneBridges*) with *cre*-specific primers that incorporated *Xba*I (5') and *Sfi*I (3') at the ends of the PCR product. This product was digested with *Xba*I and *Sfi*I and cloned into the identical sites of pCK, a *bla*$^-$, Cm^R derivative of pCK110919-HC-Bla (pACYC ori), yielding pCK-*cre*.

[354] The second feature is the naïve A domain library flanked by two orthologous *loxP* sites, *loxP*(wild-type) and *loxP*(FAS), which are required for the site-specific DNA recombination catalyzed by Cre. *See, e.g.*, Siegel, R.W., *et al.*, FEBS Letters 505:467-473 (2001). These sites rarely recombine with another. *loxP* sites were built into pCK-*cre* sequentially. 5'-phosphorylated oligonucleotides loxP(K) and loxP(K_rc), carrying loxP(WT) and *Eco*RI and *Hin*DIII-compatible overhangs to allow ligation to digested *Eco*RI

and *Hin*DIII-digested pCK, were hybridized together and ligated to pCK-*cre* in a standard ligation reaction (T4 ligase; overnight at 16°C).

[355] The resulting plasmid was digested with *EcoR*I and *Sph*I and ligated to the hybridized, 5'-phosphorylated oligos loxP(L) and loxP (L_rc), which carry *loxP*(FAS) and *EcoR*I and *Sph*I-compatible overhangs. To prepare for library construction, a large-scale purification (Qiagen MAXI prep) of pCK-*cre*-lox-P(wt)-loxP(FAS) was performed according to Qiagen's protocol. The Qiagen-purified plasmid was subjected to CsCl gradient centrifugation for further purification. This construct was then digested with *Sph*I and *Bgl*II and ligated to digested naïve A domain library insert, which was obtained via a PCR-amplification of a preexisting A domain library pool. By design, the *loxP* sites and Mb are in-frame, which generates Mbs with *loxP*-encoded linkers. This library was utilized in the *in vivo* recombination procedure as detailed below.

[356] fUSE5HA-Mb-lox-lox vector. The vector is a derivative of fUSE5 from George Smith's laboratory (University of Missouri). It was subsequently modified to carry an HA tag for immunodetection assays. *loxP* sites were built into fUSE5HA sequentially. 5'phosphorylated oligonucleotides loxP(I) *and* loxP(I)_rc, carrying loxP(WT), a string of stop codons and *Xma*I and *Sfi*I-compatible overhangs, were hybridized together and ligated to *Xma*I- and *Sfi*I-digested fUSE5HA in a standard ligation reaction (New England Biolabs T4 ligase; overnight at 16C).

[357] The resulting phage vector was next digested with *Xma*I and *Sph*I and ligated to the hybridized oligos loxP(J) and loxP(J)_rc, which carry *loxP*(FAS) and overhangs compatible with *Xma*I and *Sph*I. This construct was digested with *XmaI*/*Sfi*I and then ligated to pre-cut (*XmaI*/*Sfi*I) naïve A domain library insert (PCR product). The stop codons are located between the *loxP* sites, preventing expression of *gIII* and consequently, the production of infectious phage.

[358] The ligated vector/library was subsequently transformed into an *E. coli* host bearing a *gIII*-expressing *plasmid* that allows the rescue of the fUSE5HA-Mb-lox-lox phage, as detailed below.

[359] pCK-*gIII*. This plasmid carries *gIII* under the control of its native promoter. It was constructed by PCR-amplifying *gIII* and its promoter from VCSM13 helper phage (Stratagene) with primers gIIIPromoter_EcoRI and gIIIPromoter_HinDIII. This product was digested with *Eco*RI and *Hin*DIII and cloned into the same sites of pCK110919-HC-Bla. As *gIII* is under the control of its own promoter, *gIII* expression is presumably constitutive. pCK-*gIII* was transformed into *E. coli* EC100 (Epicentre).

[360] ***In vivo* recombination procedure.** In summary, the procedure involves the following key steps: a) Production of infective (*i.e.* rescue) of fUSE5HA-Mb-lox-lox library with an *E. coli* host expressing *gIII* from a plasmid; b) Cloning of 2nd library (pCK) and transformation into F^+ TG1 *E. coli*; c) Infection of the culture carrying the 2nd library with the rescued fUSE5HA-Mb-lox-lox phage library.

[361] *a. Rescue of phage vector.* Electrocompetent cells carrying pCK-*gIII* were prepared by a standard protocol. These cells had a transformation frequency of 4 x 10^8/μg DNA and were electroporated with large-scale ligations (~5 μg vector DNA) of fUSE5HA-lox-lox vector and the naïve A domain library insert. After individual electroporations (100 ng DNA/electroporation) with ~ 70 μL cells/cuvette, 930 μL warm SOC media were added, and the cells were allowed to recover with shaking at 37C for 1 hour. Next, tetracycline was added to a final concentration of 0.2 μg/mL, and the cells were shaken for ~ 45 minutes at 37C. An aliquot of this culture was removed, 10-fold serially diluted and plated to determine the resulting library size (1.8 x 10^7). The remaining culture was diluted into 2 x 500 mL 2xYT (with 20 μg/mL chloramphenicol and 20 μg/mL tetracycline to select for pCK-*gIII* and the fUSE5HA-based vector, respectively) and grown overnight at 30C.

[362] Rescued phage were harvested using a standard PEG/NaCl precipitation protocol. The titer was approximately 1 x 10^{12} transducing units/mL.

[363] *b. Cloning of the 2nd library and transformation* into *an E. coli host.* The ligated pCK/ naïve A domain library is electroporated into a bacterial F^+ host, with an expected library size of approximately 10^8. After an hour-long recovery period at 37C with shaking, the electroporated cells are diluted to OD_{600}~ 0.05 in 2xYT (plus 20 μg/mL chloramphenicol) and grown to mid-log phase at 37C before infection by fUSEHA-Mb-lox-lox.

[364] *c. Infection of the culture carrying the 2nd library with the rescued fUSE5HA-Mb-lox-lox phage library.* To maximize the generation of recombinants, a high infection rate (> 50%) of *E.coli* within a culture is desirable. The infectivity of *E. coli* depends on a number of factors, including the expression of the F pilus and growth conditions. *E. coli* backgrounds TG1 (carrying an F') and K91 (an Hfr strain) were hosts for the recombination system.

[365] Oligonucleotides:

loxP(K)
[P-5' agcttataacttcgtatagaaaggtatatacgaagttatagatctcgtgctgcatgcggtgcg]

loxP(K_rc)
[P-5' aattcgcaccgcatgcagcacgagatctataacttcgtatataccttttctatacgaagttataagct]

loxP(L)
[P-5' ataacttcgtatagcatacattatacgaagttatcgag]

loxP (L_rc)
[P-5' ctcgataacttcgtataatgtatgctatacgaagttatg]

loxP(I)
[P5'ccgggagcagggcatgctaagtgagtaataagtgagtaaataacttcgtatataccttttctatacgaagttatcgtctg]

loxP(I)_rc
[P-5' acgataacttcgtatagaaaggtatatacgaagttatttactcacttattactcacttagcatgccctgctc]

loxP(J)
[5' ccgggaccagtggcctctggggccataacttcgtatagcatacattatacgaagttatg]

loxP(J)_rc
[5' cataacttcgtataatgtatgctatacgaagttatggccccagaggccactggtc]

gIIIPromoter_EcoRI
[5' atggcgaattctcattgtcggcgcaactat

gIIIPromoter_HinDIII
[5' gataagctttcattaagactccttattacgcag]

Example 7

[366] This example describes optimization of multimers by optimizing monomers and/or linkers for binding to a target.

[367] Figure 8 illustrates an approach for optimizing multimer binding to targets, as exemplified with a trimeric multimer. In the figure, first a library of monomers is panned for binding to the target (*e.g.*, BAFF). However, some of the monomers may bind at locations on the target that are far away from each other, such that the domains that bind to these sites cannot be connected by a linker peptide. It is therefore useful to create and screen a large library of homo- or heterotrimers from these monomers before optimization of the monomers. These trimer libraries can be screened, *e.g.*, on phage (typical for heterotrimers created from a large pool of monomers) or made and assayed separately (*e.g.*, for homotrimers). By this method, the best trimer is identified. The assays may include binding assays to a target or agonist or antagonist potency determination of the multimer in functional protein- or cell-based assays.

[368] The monomeric domain(s) of the single best trimer are then optimized as a second step. Homomultimers are easiest to optimize, since only one domain sequence exists, though heteromultimers may also be synthesized. For homomultimers, an increase in binding by the multimer compared to the monomer is an avidity effect.

[369] After optimization of the domain sequence itself (*e.g.*, by recombining or NNK randomization) and phage panning, the improved monomers are used to construct a dimer with a linker library. Linker libraries may be formed, *e.g.*, from linkers with an NNK composition and/or variable sequence length.

[370] After panning of this linker library, the best clones (*e.g.*, determined by potency in the inhibition or other functional assay) are converted into multimers composed of multiple (*e.g.*, two, three, four, five, six, seven, eight, *etc.*) sequence-optimized domains and length- and sequence-optimized linkers.

[371] To demonstrate this method, a multimer is optimized for binding to BAFF. The BAFF binding clone, anti-BAFF 2, binds to BAFF with nearly equal affinity as a trimer or as a monomer. The linker sequences that separate the monomers within the trimer are four amino acids in length, which is unusually short. It was proposed that expansion of the linker length between monomers will allow multiple binding contacts of each monomer in the trimer, greatly enhancing the affinity of the trimer compared to the monomer molecule.

[372] To test this, libraries of linker sequences are created between two monomers, creating potentially higher affinity dimer molecules. The identified optimum linker motif is then used to create a potentially even higher affinity trimer BAFF binding molecule.

[373] These libraries consist of random codons, NNK, varying in length from 4 to 18 amino acids. The linker oligonucleotides for these libraries are:

1. 5'-AAAACTGCAATGACNNMNNMNNMNNACAGCCTGCTTCATCCGA-3'
2. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNACAGCCTGCTTCATCCGA-3'
3. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNACAGCCTGCTTCATCCGA-3'
4. 5'AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNMNNMNNACAGC CTGCTTCATCCGA-3'
5. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNACA GCCTGCTTCATCCGA-3'
6. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNN MNNMNNACAGCCTGCTTCATCCGA-3'
7. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNN NMNNMNNMNNACAGCCTGCTTCATCCGA-3'
8. 5'-AAAACTGCAATGACNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNN MNNMNNMNNMNNMNNACAGCCTGCTTCATCCGA-3'

[374] Libraries of these sequences are created by PCR. A generic primer, SfiI (5'-TCAACAGTTTCGGCCCCAGA-3'), is used with the linker oligonucleotides in a PCR with the clone anti-BAFF2 as template. The PCR products are purified with Qiagen Qiaquick columns and then digested with BsrDI. The parent anti-BAFF 2 clone is digested with BpmI. These digests are purified with Qiagen Qiaquick columns and ligated together. The ligation is amplified by 10 cycles of PCR with the SfiI primer and the primer BpmI (5'-ATGCCCCGGGTCTGGAGGCGT-3'). After purification with Qiagen Qiaquick columns, the DNAs are digested with XmaI and SfiI. Digestion products are separated on 3% agarose gel and the Dimeric BAFF domain fragments are purified from the gel. The DNA fragments are ligated into the corresponding restriction sites of phage display vector fuse5-HA, a derivative of fuse5 carrying an in-frame HA-epitope. The ligation mixture is electroporated into TransforMax™ EC100™ electrocompetent *E. coli* cells. Transformed *E. coli* cells are grown overnight at 37°C in 2xYT medium containing 20 μg/ml tetracycline. Phage particles are purified from the culture medium by PEG-precipitation and used for panning.

Example 8

[375] This example describes intra-domain recombination to identify monomer domains with improved function.

[376] Monomer sequences were generated by several steps of panning and one step of recombination to identify monomers that bind to either the CD40 ligand or human serum albumin. CD40L and HSA was panned against three different A-domain phage libraries. After two rounds of panning, the eluted phage pools were PCR amplified with two sets of oligonucleotides to produce two overlapping fragments. The two fragments were then fused together and cloned into the phagemid vector, pID, to fuse the products of two-fragment recombination. The recombined libraries (10^{10} size each) were then panned two rounds against CD40L and HSA targets using solution panning and streptavidin magnetic bead capture.

[377] The selected phagemid pools were then recloned into the protein expression vector, pET, a T7 polymerase driven vector, for high protein expression. Almost 1400 clones were screened for anti-CD40L binding monomers by standard ELISA and about 2000 clones were screened for HSA. All clones were unique sequences.

[378] ELISA plate wells were coated with 0.2 μg of CD40L or 0.5 μg of HAS, and 5 μl of the monomer expression clone lysate was applied to each well. The bound monomers (which were produced as a hemagglutinin (HA) fusion) were then detected by

anti-HA-HRP conjugated antibody, developed by horse-radish peroxidase enzyme activity, and read at an OD of 450 nm. The positive clones were selected by comparing the ELISA reading to the existing trimer anti-CD40L 2.2 and were selected and sequenced with the T7 primer.

[379] For the anti-CD40L samples, two anti-CD40L 2.2Ig clones were grown in the same plate with selected monomer clones and processed side by side as the positive control. Two empty pET vector clones transformed were grown and processed as negative controls. The ELISA reading at OD450 and the corresponding clone sequences are shown.

[380] The same selection and screen processes apply to HSA. Existing anti-HSA monomer and trimer were used as positive controls, empty pET vector were used as negative controls. Positive binders were selected as those with an ELISA signal equal or better than the anti-HSA trimer.

[381] The positive rate of clones with an OD_{450} greater or equal to the anti-CD40L2.2Ig binding was about 0.7% for CD40L and 0.4% for HSA.

[382] Identified sequences are listed below:

Anti-CD40L positive clones after 2 fragments recombination and solution panning

pmA2_84	CRPNQFT	CGNGH	CLPRTWL	CDGVPD	CQDSSDETPIP	CKSSVPTSLQ
A5C1	CQSSQFR	CRDNST	CLPLRLR	CDGVND	CRDGSDESPAL	CGRPGPGATSAPAASLQ
pmA2_18	CPADQFQ	CKNGS	CIPRPLR	CDGVED	CADGSDEGQD	CGRPGPGATSAPAASLQ
pmA5_79	CARDGEFR	CAMNGR	CIPSSWV	CDGEDD	CGDGSDESQVY	CGGGGSLQ
A2F10	CLPSQFP	CQNSSI	CVPPALV	CDGDAD	CGDDSDEAS	CAPPGSLSLQ
A1E9	CAPGEFT	CGNGH	CLSRALR	CDGDDG	CLDNSDEKN	CPQRTSLQ
pmA11_40	CLANECT	CDSGR	CLPLPLV	CDGVPD	CEDDSDEKN	CTKPTSLQ

Anti-HSA positive clones after 2 fragments recombination and solution panning

A5B_10	CRPSQFR	CGSGK	CIPQPWG	CDGVPD	CEDNSDETD	CKTPVRTSLQ
A5_2_68	CPASQFR	CENGH	CVPPEWL	CDGVDD	CQDDSDESSAT	CQPRTSLQ
A5_8_93	CAPGQFR	CRNYGT	CISLRWG	CDGVND	CGDGSDEQN	CTPHTSLQ
A1_4	CLANQFK	CESGH	CLPPALV	CDGVDD	CQDSSDEASAN	C
A1_34	CNPTGKFK	CRSGR	CVPRESCR	CDGVDD	CEDNSDEKD	CQPHTSLQ
A2_10	CESSEFQ	CENGH	CLPVPWL	CDGVND	CADGSDEKN	CPKPTSLQ

[383] While this example demonstrates the use of LDL-receptor A domains, those of skill in the art will appreciate that the same techniques can be used to generate desired binding properties in monomer domains of the present invention.

Example 9

[384] This example describes an exemplary method for the design and analysis of libraries comprising monomers that comprise only residues observed in natural domains at any given sequence position. To this end, a sequence alignment of all natural domains of a given family is constructed. Since the cysteine residues tend to be the most conserved feature of the alignment, these residues are used as a guide for further design. Each stretch of sequence between two cysteines is considered separately to account for structural variability due to length variations. For each inter-cysteine sequence, a histogram of lengths is constructed. Lengths observed at roughly 10% or greater frequency in known domains are considered for use in the library design. A separate alignment of sequences is constructed for each length, and amino acids which occur at greater than approximately 5% at a given position in the sub-alignment are allowed in the final library design for that length. This process is repeated for each inter-cysteine sequence segment to generate the final library design. Oligonucleotides with degenerate codons designed to optimally express the desired protein diversity are then synthesized and assembled using standard methods to create the final library.

[385] Typically four sets of overlapping oligonucleotides are designed with a 9-base overlap between sets 1 and 2, sets 2 and 3, as well as sets 3 and 4 for PCR assembly. In some cases, two sets of overlapping oligonucleotides are designed with a 9-base overlap between the two sets. The libraries are constructed with the following protocol:

[386] *Oligonuleotides*: A 10 μM working solution of each oligonucleotide is prepared. Equal molar amounts of oligos for each set are mixed (sets 1, 2, 3 and 4). The oligonucleotides are assembled in two PCR assembly steps: the first round of PCR assembles sets 1 and 2, as well as sets 3 and 4 and the the second round of PCR uses the first round PCR products to assemble the full length of each library.

[387] *PCR assembly - Round 1*: Separate PCR reactions are performed done using the following pairs of oligos: each oligo from set 1 vs. pooled set 2; each oligo from set 2 vs. pooled set 1; each oligo from set 3 vs. pooled set 4; each oligo from set 4 vs. pooled set 3. PCR reaction mixtures are 50 μL in volume and comprise 5 μL 10X PCR buffer, 8 μL 2.5 mM dNTPs, 5 μL each of oligo and its pairing oligo pool, 0.5 μL LA Taq polymerase and 26.5 μL water. PCR reaction conditions are as follows: 18 cycles of [94°C/10", 25 °C/30", 72 °C/30"] and 2 cycles of [94°C/30", 25°C/30", 72°C /1']. 5 μL of each PCR reaction is run

on 3% low-melting Agrose gel in TBE buffer to verify the presence of expected PCR product.

[388] *PCR assembly - Round 2*: All Round 1 PCR products are pooled with 5 μL from each PCR reaction. The full length product of each library scaffold is assembled by PCR using a reaction volume of 50 μL comprising 4 μL 10X PCR buffer, 8 μL 2.5 mM dNTPs, 10 μL pooled Round 1 PCR products, 0.5 μL LA Taq and 27.5 μL water and the following reaction conditions: 8 cycles of [94°C/10", 25°C/30", 72°C/30"] and 2 cycles of [94°C/30", 25°C/30", 72°C/1'].

[389] *Rescue PCR and Sfi digestion:* The fully assembled library scaffolds are amplified via PCR to generate sufficient material for library production. Four separate 50 μL- PCR reactions are performed. Each reaction mixture comprises: 2.5 μL 10X PCR buffer, 8 μL 2.5 mM dNTPs, 25 μL Round-2 PCR products, 0.5 μL LA Taq, 5 μL each of 10 μM 5' and 3' Rescue PCR primers (Table 2), and 4 μL water. The reaction conditions are as follows: 8 cycles of [94°C/10", 25°C/30", 72°C/30"] and 2 cycles of [94°C/30", 45°C/30", 72°C/1']. 5 μL of the reaction mixture is run on a 3% low-melting Agrose gel in TBE buffer to confirm that the amplification product is the correct size. The amplification product is then purified by QIAGEN QIAquick columns, eluted in EB buffer, and digested with *Sfi* restriction enzyme for cloning to *Sfi*-digested ARI 2 vector. Twenty μg of the assembled library scaffold is digested with 200 units of *Sfi* restriction enzyme in 1,000 μL total volume and 3 hrs at 50°C. The digested DNA is purified with QIAGEN QIAquick columns and eluted in water.

[390] *Test ligation:* To determine the optimal library insert/vector ratio for ligation, 1 μL of each a dilution series of *Sfi*-digested library insert (1/1, 1/5, 1/25, 1/125 and 1/625) is used for ligation with 1 μL *Sfi*-digested ARI 2 vector, 1 μL T4 DNA ligase, 1 μL 10X ligase buffer and 7 μL water. The ligation reaction mixture is incubated at room temperature for 2 hours to generate a ligated product. 1 μL ligated product is mixed with 40 μL EC100 cells in 0.1 cm cuvette, incubated on ice for 5 minutes, electroporated, and recovered in 1 mL SOC for 1 hour at 37°C. For each electroporation, 5 μL each of dilution series (1/1, 1/10, 1/100, 1/1,000) is spotted on Agar plate with Tetracycline to determine the optimal inert/vector ratio. In addition, 50 μL of each of dilution is plated to grow single colonies for library QC.

[391] *Sequence Analysis and Protein Expression:* Individual clones are picked and grown overnight in 0.4 mL 2xYT with 20 μg/mL tetracycline in 96-well plates. The overnight grown cells are spun down, and 0.5 μL 1/5 dilute supernatant is used to amplify the library inserts using 5' and 3' rescue primer for sequencing. DNA sequence analyses is used to verify the presence of the expected library inserts. To examine the protein expression, the library inserts are transferred to a pEVE expression vector. The 0.5 μL of pooled supernatants of selected clones from overnight-culture are amplified using a pair of PCR primers with Sfi restriction sites that are in-frame with HA epitope at the N-terminus and His8 Tag at the C-terminus. The PCR reaction mixture comprises: 0.5 μL phage (pool of 32 supernatants), 5 μL 10x LA Taq buffer, 8 μL 2.5 mM dNTPs, 5 μL each of 10 μM EGF Eve 5 and 10 μM 3Sfi N primers, and 0.5 μL LA Taq polymerase. The PCR reaction conditions are as follows: 23 cycles of [94°C/10", 45°C/30", 72°C/30"] and 2 cycles of [94°C/", 45°C/30", 72°C/1']. The amplification product is purified by QIAquick columns and digested with *Sfi* enzyme, and ligated with *Sfi*-digested pEVE vector for 2 hours at room temperature according to manufacture's specifications. 1 μL of the ligated product is transformed in 40 μL BL21 cells by electroporation, plated on Kanamycin plate, and grown in the 37°C incubator overnight. Colonies are picked and cultured overnight in 0.5 mL 2xYT media. The following day, 50 μL of overnight culture is inoculated to 1 mL 2xYT media and grown for about 2.5 hours until OD600 reached about 0.8, at which point IPTG is added to a final concentration of 1 mM for protein expression. The cells are spun down at 3,600 rpm for 15 minutes, the pellets are suspended in 100 μL TBS/2 mM Ca^{++}, heated at 65°C for 5 minutes to release the protein, and spun down at 3,600 rpm for 15 minutes. The supernatant from each clone is run on a 4-12% NuPAGE gel, 10 μL each with or without reducing agent (Invitrogen). Shift in band position between reduced and unreduced samples indicates that the expressed proteins are likely to fold properly.

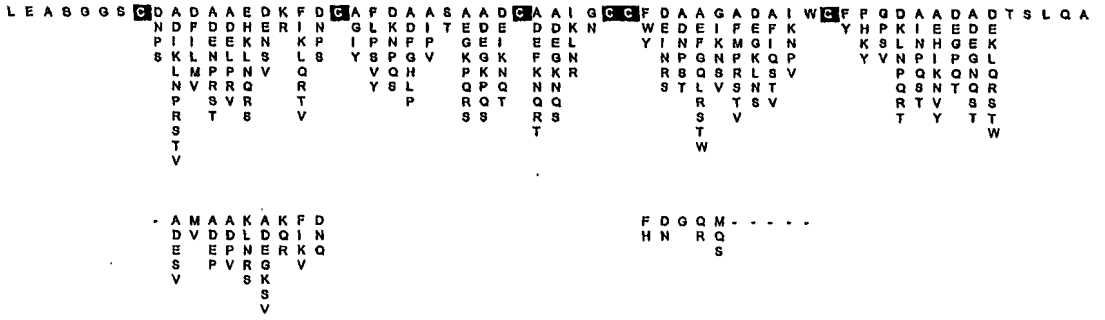
[392] *Library Scale-up*: The full library is ligated in a ARI 2 vector, transformed in EC100 cells, then expanded in K91 cells. The ligation is performed overnight at room temperature in a final volume of 2.5 mL with 25 μg of *Sfi*-digested vector, 2.5 μg *Sfi*-digested library insert, 5 μL T4 DNA ligase, and 250 μL 10x DNA ligase buffer. The ligated product is precipitated with sodium acetate and ethanol, suspended in 400 μL water, reprecipitated with NaAc/EtOH and resuspended in 50 μL H2O. The library is electroporated in a vessel comprising 10 μL DNA and 200 μL EC100 cells, transferred to 50 mL SOC

media, and grown at 37°C for 1 hour at 300 rpm. A 5 μL aliquot is removed and (1) serially diluted to determine the library size; and (2) plated out for sequence verification. The transformed EC100 in 50 mL SOC is divided equally, added to six 500 mL culture of K91 cells with OD600 of 0.5, and incubated for 30 minutes at 37 C without shaking. Tetracycline is added to a concentration 0.2 μg/mL, and the cultures are grown for 30 minutes at 37°C at 300 rpm. Finally, tetracycline is added to a final concentration 20 μg/mL, and the cultures are grown overnight at 37°C at 300 rpm. Cells are centrifuged at 8,000 rpm for 10 minutes. Phages in the supernatant are precipitated by adding 40 g PEG and 30 g NaCl /1000 mL, and centrifugation at 8,000 rpm for 10 minutes. Phages are resuspended in 50 mL TBS/2 mM Ca^{++} and centrifuged at 5,000 rpm for 10 minutes to remove the cell debris. The supernatant is added with a final concentration of 20% PEG and 1.5 M NaCl, and placed on ice for 40 minutes, and phages are spun down at 5,000 rpm for 10 minutes, and resuspended in 10 mL TBS/2 mM Ca^{++}. Phage titer is determined by serial dilution.

Example 10

[393] This example describes design and analysis of a library from trefoil/PD domains using the methods set forth in Example 9 above.

[394] Based on sequence alignments of naturally occurring trefoil/PD domains, a panel of degenerate oligonucleotides were designed that encode trefoil/PD domains that comprise amino acids at each position that are found only in naturally occurring trefoil/PD domains. The trefoil/PD library design is set forth below.



[395] The degenerate oligonucleotide sequences are set forth in the table below:

PD1_1_1	CTG GAG GCG TCT GGT GGT TCG TGT YCN SYA WTK RAY GWB MRY GWS ARR AVA GAC TGC GCG
PD1_1_2	CTG GAG GCG TCT GGT GGT TCG TGT RAY ANM GWY MSY CBN CWR ARY ARR CWA GAC TGC GCG
PD1_1_3	CTG GAG GCG TCT GGT GGT TCG TGT RAY ANM WTK GMR CBN RAR GWS ARR DTC GAC TGC GCG
PD1_1_4	CTG GAG GCG TCT GGT GGT TCG TGT RAY SYA GWY GMR GWB RAR ARY ARR DTC GAC TGC GCG
PD1_2_1	CTG GAG GCG TCT GGT GGT TCG TGT TCN RTG SCN GWY CTN KCN MRR AWA GAC TGC GCG
PD1_2_2	CTG GAG GCG TCT GGT GGT TCG TGT GVS RTG GAD SCN ARN GDY MRR KTY GAC TGC GCG

PD1_2_3	CTG GAG GCG TCT GGT GGT TCG TGT GVS RTG SCN SCN CTN RAR MRR KTY GAC TGC GCG
PD1_2_4	CTG GAG GCG TCT GGT GGT TCG TGT TCN RTG GAD GWY ARN RAR MRR AWA GAC TGC GCG
PD2_1	GCA GCA CCC TMK YTB RAA RCA WRT YYB YYB RST DAY AAR NGR DGR CGC GCA GTC
PD2_2	GCA GCA CCC NTT BGC YYG RCA YTB NGR CGV RST NGS RBC RTY RWA CGC GCA GTC
PD2_3	GCA GCA CCC NTT RYY WKY RCA RTY RBC YYB RST NGS RKG YTK YAM CGC GCA GTC
PD2_4	GCA GCA CCC TMK RYY WKY RCA RTY RBC CGV RST DAY RKG YTK YAM CGC GCA GTC
PD3_1_1	GGG TGC TGC TWY MGY HCN DSG RKY KYY RAR DYY AAH TGG TGC TAC
PD3_1_2	GGG TGC TGC TGG AWY RMY SAR AAH ABG YTR CAR RTH TGG TGC TAC
PD3_1_3	GGG TGC TGC TWY GAS RMY YTT RKY BCN RRY CAR CCN TGG TGC TAC
PD3_1_4	GGG TGC TGC TWY GAS HCN YTT AAH BCN RRY DYY RTH TGG TGC TAC
PD3_2_1	GGG TGC TGC TTY RAY GGA CRR ATG TGG TGC TAC
PD3_2_2	GGG TGC TGC AAY RAY GGA CRR CAR TGG TGC TAC
PD3_2_3	GGG TGC TGC AAY RAY GGA CRR TCN TGG TGC TAC
PD3_2_4	GGG TGC TGC TTY RAY GGA CRR TCN TGG TGC TAC
PD4_1	GGC CTG CAA TGA CGT CSW RBY NGK RTD YKG YMG NGR YTT GTA GCA CCA
PD4_2	GGC CTG CAA TGA CGT YWK YTS YTS YDC RHT RTY NMC RAA GTA GCA CCA
PD4_3	GGC CTG CAA TGA CGT STY YTS RYC TWT NGY YKK NGR RTR GTA GCA CCA
PD4_4	GGC CTG CAA TGA CGT STY RBY RYC TWT NGY YKK NMC RTR GTA GCA CCA
5'	Rescue 5'_AAAAGGCCTCGAGGGCCTGGAGGCGTCTGGTGGTTCGTGT_3'
3'	Rescue 5'_AAAAGGCCCCAGAGGCCTGCAATGACGT_3'

[396] N represesents A, T, G, or C: B represents G, C, or T; D represents G, A, or T; H represents A, T, or C; K represents G or T; M represents A or C; R represents A or G; S represents G or C; V represents G, A, or C; W represents A or T; and Y represents T or C.

[397] Thirty two individual phages from each library were amplified by PCR and the amplification products were sequenced. The results of sequencing confirmed that the phage contained inserts of the expected sizes and sequences for the library. The library comprised 2.31×10^9 monomer domains comprising 57, 58, 61, or 62 amino acids. The sequencing results are shown in the table below. Clones 5 and 6 were identified as clones that do not contain a domain insert, but instead represent empty vector background from the transformation.

PD_1	PGLEGLEASGGSCDANEVKNKFDCAYDAATPSQCRAKGCCWINQNTLQIWCYFGNNEEEQTSLQASGA
PD_2	PGLEGLEASGGSCDIDSRLNKQDCAVKPPSEGDCENNGCCFNGQMWCYFGNSEKKKTSLQASGA
PD_3	PGLEGLEASGGSCGVEPNGQVDCAFDGPTSSKCQANGCCNNGRS*CYFVNNAKQKTSLQASGA
PD_4	PGLEGLEASGGSCDMEAKGRVDCAFNGASASECRANGCCNNGQQWCYKSRPYTASTSLQASGA
PD_5	PGLEGH**LCYEASGA
PD_6	PGLEGH**LCYEASGA
PD_7	PGLEGLEASGGSCAVPALKRFDCALKPVSPADCAGRGCCNNGQQWCYKSLQYTGSTSLQASGA
PD_8	PGLEGLEASGGSCNRDRLLNRLDCAYDAASPPKCRANGCCFNGQMWCYYPPTIGEDTSLQASGA
PD_9	PGLEGLEASGGSCDNLAREVKIDCAVKHASETDCDNNGCCWNDENRLQVWCYFGNSEQKKTSLQASGA
PD_10	PGLEGLEASGGSCSMAVLAQKDCAVQHPTKADCENKGCCNNGRSWCYKPLQNTNWTSLQASGA
PD_12	PGLEGLEASGGSCAVAPLERFDCALQHATRADCANKGCCFGQMWCYKSRQNPDTTSLQASGA
PD_13	PGLEGLEASGGSCGVEPKGKVDCAPPLVSEQTCFKRGCCFDGQMWCYYGKTKDNNTSLQASGA
PD_15	PGLEGLEASGGSCDAVEKENKFDCAVQHASRANCENNGCCNNGQSWCYHVTAKDANTSLQASGA
PD_16	PGLEGLEASGGSCSVPDLAKKDCALKPITAANCEDIGCCFDGRQWCYFGDNAEQKTSLQASGA
PD_17	PGLEGLEASGGSCPPINEHERRDCAVKHATKADCDGNGCCFDDLGADQPWCYFVDNAEKKTSLQASGA
PD_19	PGLEGLEASGGSCSVPVLSKIDCAVKHPSRANCENNGCCNNGQSWCYYVQTKGNKTSLQASGA
PD_20	PGLEGLEASGGSCDKDSPLSKLDCAPSLITRRTCFELGCCNNGRQWCYFGNNAEQITSLQASGA

PD_21	PGLEGLEASGGSCEVPALEKFDCAYDDPSAPKCQAKGCCFNGQMWCYYGKTKDTDTSLQASGA
PD_22	PGLEGLEASGGSCDMEAKVRFDCAVQHPTRDNCDSKGCCNNGQSWCYFGNNAQQKTSLQASGA
PD_23	PGLEGLEASGGSCGVAALEQFDCALKHPSGDNCDSNGCCFDGRMWCYHSQTKGQETSLQASGA
PD_25	PGLEGLEASGGSCSAINVSVRTDCAVKHVSPGDCNDLGCCNNGQSWCYHVPAIGNETSLQASGA
PD_27	PGLEGLEASGGSCAMPPLEQFDCAVKPITADDCANRGCCFNGQMWCYYPPTINEDTSLQASGA
PD_29	PGLEGLEASGGSCGMEARVKVDCAYDDATPPKCQANGCCNNGQSWCYFGNNAQQQTSLQASGA
PD_30	PGLEGLEASGGSCGVAALERVDCAVKHPTEGNCTSNGCCFDGQMWCYKPRQNTDSTSLQASGA
PD_31	PGLEGLEASGGSCDVEANGQVDCALKHATGNDCASNGCCFDGQSWCYHPKAINENTSLQASGA
PD_32	PGLEGLEASGGSCDANENESKVDCALQHVTSGDCTDIGCCFNGQSWCYYVQAIGANTSLQASGA

[398] Clones from the trefoil/PD library were tested for their ability to produce folded protein. SDS-PAGE verified that the clones produced full-length soluble protein following heat lysis.

Example 11

[399] This example describes design and analysis of a library from thrombospondin domains using the methods set forth in Example 9 above.

[400] Based on sequence alignments of naturally occurring thrombospondin domains, a panel of degenerate oligonucleotides were designed that encode thrombospondin domains that comprise amino acids at each position that are found only in naturally occurring thrombospondin domains. The thrombospondin library design is set forth below.

[401] The degenerate oligonucleotide sequences are set forth in the table below:

T1_1	CTG GAG GCG TCT GGT GGT TCG TGT AVY RSH GMN TGT GRN ARY GGT WBB RTH WHY DCN BMY CKN GGC TGC GAC
T1_2	CTG GAG GCG TCT GGT GGT TCG TGT AVY VDA AVY TGT KCN VNG GGT VAR WCN RWG CRR SWA RYG GGC TGC GAC
T1_3	CTG GAG GCG TCT GGT GGT TCG TGT AVY VDA CVR TGT KCN ARY GGT YWY MRR CRS CRR ANA RYG GGC TGC GAC

T1_4	CTG GAG GCG TCT GGT GGT TCG TGT AVY RSH CVR TGT KCN VNG GGT YWY MRR CRS CRR ANA CKN GGC TGC GAC
T2_1_1	CTC CGG GCA NGD BGM NCC NGR RKS NCC HSC NSC GTC GCA GCC
T2_1_2	CTC CGG GCA RBC NCB YRA YYC RAA RAA YDG YKK GTC GCA GCC
T2_1_3	CTC CGG GCA RBC NCB YRA YYC YST YWG YGA YKK GTC GCA GCC
T2_2_1	CTC CGG GCA CAY NYC NSC RKC YTS YCS RTT YKC RTC GTC GCA GCC
T2_2_2	CTC CGG GCA YYC NGW RCT YRR RKC YDG YYG MTG NGA GTC GCA GCC
T2_2_3	CTC CGG GCA CVG NGW RCT YRR NGT YDG RAA CRT YTT GTC GCA GCC
T2_3_1	CTC CGG GCA RWW YYK NCC NCC YCS YRA NCC YKY YKG YTG GTC GCA GCC
T2_3_2	CTC CGG GCA YKS RYT NCC RTT RTK HSC NGS RBK NAC RHK GTC GCA GCC
T2_3_3	CTC CGG GCA NTM NGC NCC RTT RWA YYK NGS YRM NAC NGC GTC GCA GCC
T2_4_1	CTC CGG GCA RAA RTC RRA YKS YRM DAY HYS NSC RBY RTB YKT GTC GCA GCC
T2_4_2	CTC CGG GCA YTS YTS YYC RYT YRM RSK YGW RTT YYG NGV RYB GTC GCA GCC
T2_4_3	CTC CGG GCA RYT NGW RTS RTY YRM YTS YGW RAM RWW YTT RAA GTC GCA GCC
T2_5_1	CTC CGG GCA VWR YYT YTC BTC NAS KGH KMT YTC YGT NSC RRM NGV YTT NGG YCK GTC GCA GCC
T2_5_2	CTC CGG GCA NGC RTC RTS RDG NAS KGH YTY YAR YTY YTG YKS NGV YAR YTT YTB GTC GCA GCC
T2_5_3	CTC CGG GCA YYT NGA NGR RCT NAS KGH YTY YAR YTY YTG RKY NGV YAR YTT YTB GTC GCA GCC
T3_1_1	TGC CCG GAG CNR CKN GHR GAN THY CRR RAK TGT WMY MBG VAN GCC TGC GGC
T3_1_2	TGC CCG GAG GMY GWR AVR CRR RHA ATA KYR TGT SRN SMR SVK GCC TGC GGC
T3_1_3	TGC CCG GAG AVY RVY TYW CRR RHA RMR MSS TGT SRN RNY SVK GCC TGC GGC
T3_2_1	TGC CCG GAG SAR GYN ARR CCG SMR GMN CDR VAR CVR TGT WMY MBG VAN GCC TGC GGC
T3_2_2	TGC CCG GAG KCN WCN ARR CCG ARY NCN RMR AGB DCN TGT SRN SMR SVK GCC TGC GGC
T3_2_3	TGC CCG GAG CWY CHR ARR CCG ARY ATY RMR AGB DCN TGT SRN RNY SVK GCC TGC GGC
T4_1	GGC CTG CAA TGA CGT YKK HTC CCA YDG RBT CCA BWS GCC GCA GGC
T4_2	GGC CTG CAA TGA CGT YRM RSY RAA NKY YBC RTA RKN GCC GCA GGC
T4_3	GGC CTG CAA TGA CGT HTC HTC RAA YDG YBC SRY BWS GCC GCA GGC
T4_4	GGC CTG CAA TGA CGT YKK RSY CCA NKY RBT SRY RKN GCC GCA GGC
5'	Rescue 5'_AAAAGGCCTCGAGGGCCTGGAGGCGTCTGGTGGTTCGTGT_3'
3'	Rescue 5'_AAAAGGCCCCAGAGGCCTGCAATGACGT_3'

[402] N represesents A, T, G, or C: B represents G, C, or T; D represents G, A, or T; H represents A, T, or C; K represents G or T; M represents A or C; R represents A or G; S represents G or C; V represents G, A, or C; W represents A or T; and Y represents T or C.

[403] Thirty two individual phages from the library were amplified by PCR and the amplification products were sequenced. The results of sequencing confirmed that the phage contained inserts of the expected sizes and sequences for the library. The library comprised 1.98 x 10^9 monomer domains comprising 60-70 amino acids. The sequencing results are shown in the table below. Clones 1, 4, 8, 11, 12, 22, 26, and 30 were identified as clones that do not contain a domain insert, but instead represent empty vector background from the transformation.

Tsp1_1	PGLEGH**LCYEASGA
Tsp1_2	PGLEGLEASGGSCNDPCSRRYQQQNSGCYHENRQAGDMCPETSFXTKTCRVGACGQWNPWDTTSLQASGA
Tsp1_3	PGLEGLEASGGSCTSECDNGSVYSYLGCDFKIFSQSNDSSCPESDLRKKTCRVRACGHWSLWETTSLQASGA
Tsp1_4	PGLEGH**LCYEASGA
Tsp1_5	PGLEGLEASGGSCNGSCSVGESERVMGCDPSQTESSDCPENNSQETRCGGAACGHTNTWTQTSLQASGA
Tsp1_6	PGLEGLEASGGSCTESCSAGQSVRQMGCDDENRQAADMCPESAFRTTSCGIQACGLWNQWEQTSLQASGA
Tsp1_7	PGLEGLEASGGSCSTQCSRGHQRQRLGCDPSQRESRGCPEQLADSRKCTPEACGNYETFGSTSLQASGA
Tsp1_8	PGLEGH**LCYEASGA
Tsp1_9	PGLEGLEASGGSCNSPCARGYRHQTLGCDKTFQTLSSPCPENSFQETRCDDGACGTMSNWAPTSLQASGA
Tsp1_10	PGLEGLEASGGSCGGAACGQVPPFEETSLQASGA
Tsp1_11	PGLEGH**LCYEASGA
Tsp1_12	PGLEGH**LCYEASGA

Tsp1_13	PGLEGLEASGGSCSRSCSLGKSERETGCDDANRQDGKMCPERLEEFRKCNRKACGVPEPFEETSLQASGA
Tsp1_14	PGLEGLEASGGSCTTQCAMGYRRRKLGCDLVTAGHNGNECPELLKPNIASACDVRPCGPYATFXLTSLHASGA
Tsp1_15	PGLEGLEASGGSCSGPCAMGLQRQTLGCDDENRQAANMCPESNLRVKRCHVAACGTYEKFAATSLQASGA
Tsp1_16	PGLEGLEASGGSCTGPCAMGLKRQILGCDKLFFGSRACPEHLRPSIARTCGGGACGAYGTFTATSLQASGA
Tsp1_17	PGLEGLEASGGSCSXNCSLGKSERLAGCDQKLPEQKLETVHHDACPESGFREKRXDVGACGHYXKFCFDVIAGIWG
Tsp1_18	PGLEGLEASGGSCSIRCSKGYRHQILGCDKTFQTLSTPCPEEARPAAREPCYRKACGPATTWTQTSLQASGA
Tsp1_19	PGLEGLEASGGSCSKNCSTGQSMRQVGCDAAGDPGSSCPESGSRVKRCGSPACGLTEQFEKTSLQASGA
Tsp1_20	PGLEGLEASGGSCSKRCAPGHRRRTLGCDDENREDADMCPEEARPPDLQRCSRKACGQVEPFXKTSLQASGA
Tsp1_21	PGLEGLEASGGSCSVSCSLGESVREMGCDKTFLTLSSLCPESGFQTKRCGDRACGATNNWTPTSLQASGA
Tsp1_22	PGLEGH**LCYEASGA
Tsp1_23	PGLEGLEASGGSCSGRCAKGYRRQKRGCDPQFFELRACPEEARPAEQEPCSMDACGDVNTWAKTSLQASGA
Tsp1_24	PGLEGLEASGGSCSGTCAVGESERQMGCDSVNAGNKGSECPESNFRVKRCRGAACGPYETFTSTSLQASGA
Tsp1_25	PGLEGLEASGGSCTKNCSGGETKRQTGCDEANREDAEMCRENNSRPEMCGIGACGACGGRGPHLIAX
Tsp1_26	PGLEGH**LCYEASGA
Tsp1_27	PGLEGLEASGGSCNPNCAGGKTLQLMSCYPPFFDSRACPESDLQVXPCHGGLXWRXSRXXWGX
Tsp1_28	PGLEGLEASGGSCSGPCAKGLQRRKLGCDNSNREXAEMCPELLRPNIKRTCGNGACYQWXQWEQTSLQASGA
Tsp1_29	PGLEGLEASGGSCNVTCATGESKRVMGCDQPTGSGGGKICPESDLQIEPCRVGACGDVNAWTKTSLQASGA
Tsp1_30	PGLEGH**LCYEASGA
Tsp1_31	PGLEGLEASGGSCSTQCAMGYRQRKRGCDTSQTESRGCPENALRKTPCRTGAYGNANNWTPTSLQASGA
Tsp1_32	PGLEGLEASGGSCTGPCSMGFKRQILGCDFAYMNNANCPEXXEPADPNRCNARACGHSNACSHTSLQASGA

[404] Clones from the thrombospondin library were tested for their ability to produce folded protein. SDS-PAGE verified that the clones produced full-length soluble protein following heat lysis.

Example 12

[405] This example describes an exemplary method of generating libraries comprised of proteins with randomized inter-cysteine loops. In this example, in contrast to the separate loop, separate library approach described above, multiple intercysteine loops are randomized simultaneously in the same library.

[406] An A domain NNK library encoding a protein domain of 39-45 amino acids having the following pattern was constructed:

C1-X(4,6)-E1-F-R1-C2-A-X(2,4)-G1-R2-C3-I-P-S1-S2-W-V-C4-D1-G2-E2-D2-D3-C5-G3-D4-G4-S3-D5-E3-X(4,6)-C6;
where,
C1-C6: cysteines;
X(n): sequence of n amino acids with any residue at each position;
E1-E3: glutamine;
F: phenylalanine;
R1-R2: arginine;
A: alanine;
G1-G4: glycine;
I: isoleucine;
P: proline;
S1-S3: serine;
W: tryptophan;
V: valine;

D1-D5: aspartic acid; and
C1-C3, C2-C5 & C4-C6 form disulfides.

[407] The library was constructed by creating a library of DNA sequences, containing tyrosine codons (TAT) or variable non-conserved codons (NNK), by assembly PCR as described in Stemmer *et al.*, *Gene* 164:49-53 (1995). Compared to the native A-domain scaffold and the design that was used to construct library A1 (described previously) this approach: 1) keeps more of the existing residues in place instead of randomizing these potentially critical residues, and 2) inserts a string of amino acids of variable length of all 20 amino acids (NNK codon), such that the average number of inter-cysteine residues is extended beyond that of the natural A domain or the A1 library. The rate of tyrosine residues was increased by including tyrosine codons in the oligonucleotides, because tyrosines were found to be overrepresented in antibody binding sites, presumably because of the large number of different contacts that tyrosine can make. The oligonucleotides used in this PCR reaction are:

1. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKNNKNNKGAATTCCGA- 3′
2. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKNNKNNKNNKGAATTCCGA- 3′
3. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKNNKNNKNNKNNKGAATTCCGA-3′
4. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTTATNNKNNKNNKGAATTCCGA- 3′
5. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKTATNNKNNKNNKGAATTCCGA- 3′
6. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKTATNNKNNKGAATTCCGA- 3′
7. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKTATNNKGAATTCCGA- 3′
8. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKNNKTATGAATTCCGA- 3′
9. 5′ -ATATCCCGGGTCTGGAGGCGTCTGGTGGTTCGTGTNNKNNKNNKTATNNKGAATTCCGA- 3′
10. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNMNNTGCACATCGGAATTC- 3′
11. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNMNNMNNTGCACATCGGAATTC- 3′
12. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNMNNMNNMNNTGCACATCGGAATTC- 3′
13. 5′ -ATACCCAAGAAGACGGTATACATCGTCCATAMNNMNNTGCACATCGGAATTC- 3′
14. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNATAMNNMNNTGCACATCGGAATTC- 3′
15. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNATAMNNTGCACATCGGAATTC- 3′
16. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNMNNATATGCACATCGGAATTC- 3′
17. 5′ -ATACCCAAGAAGACGGTATACATCGTCCMNNMNNATAMNNTGCACATCGGAATTC- 3′
18. 5′ -ACCGTCTTCTTGGGTATGTGACGGGGAGGACGATTGTGGTGACGGATCTGACGAG- 3′
19. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNMNNMNNCTCGTCAG ATCCGT- 3′
20. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNMNNMNNMNNCTCGTCA GATCCGT- 3′
21. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNMNNMNNMNNMNNC TCGTCAGATCCGT- 3′
22. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAATAMNNMNNMNNCTCGTC AGATCCGT- 3′
23. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNATAMNNMNNMNNCT CGTCAGATCCGT- 3′
24. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNATAMNNMNNCTCGT CAGATCCGT- 3′
25. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNATAMNNCTCG TCAGATCCGT- 3′
26. 5′ -ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNMNNATACTCG TCAGATCCGT- 3′

27. 5' -
ATATGGCCCCAGAGGCCTGCAATGATCCACCGCCCCCACAMNNMNNMNNATAMNNCTCGTCAGATCCGT- 3'
where R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, and N=A/C/G/T

[408] The library was constructed though an initial round of 10 cycles of PCR amplification using a mixture of 4 pools of oligonucleotides, each pool containing 400pmols of DNA. Pool 1 contained oligonucleotides 1-9, pool 2 contained 10-17, pool 3 contained only 18 and pool 4 contained 19-27. The fully assembled library was obtained through an additional 8 cycles of PCR using pool 1 and 4. The library fragments were digested with XmaI and SfiI. The DNA fragments were ligated into the corresponding restriction sites of phage display vector fuse5-HA, a derivative of fuse5 carrying an in-frame HA-epitope. The ligation mixture was electroporated into TransforMax™ EC100™ electrocompetent *E. coli* cells resulting in a library of $2X10^9$ individual clones. Transformed *E. coli* cells were grown overnight at 37°C in 2xYT medium containing 20 μg/ml tetracycline. Phage particles were purified from the culture medium by PEG-precipitation and a titer of $1.1X10^{13}$/ml was determined. Sequences of 24 clones were determined and were consistent with the expectations of the library design.

[409] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques, methods, compositions, apparatus and systems described above can be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.

WHAT IS CLAIMED IS:

1. A method for identifying a monomer domain that binds to a target molecule, the method comprising,

a) providing a library of non-naturally-occurring monomer domains, wherein the monomer domain is selected from the group consisting of: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain,

wherein the thrombospondin monomer domain comprises the following sequence:

(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6;

the trefoil monomer domain comprises the following sequence:

C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and

the thyroglobulin monomer domain comprises the following sequence:

C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6;

wherein "x" is any amino acid;

b) screening the library of monomer domains for affinity to a first target molecule; and

c) identifying at least one monomer domain that binds to at least one target molecule.

2. The method of claim 1, wherein the at least one monomer domain specifically binds to a target molecule not bound by a naturally-occurring monomer domain at least 90% identical to the non-naturally occurring monomer domain,

3. The method of claim 1, wherein

C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds; and

C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds.

4. The method of claim 1, wherein

the thrombospondin monomer domain comprises the following sequence:

(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]xC_3[vldr]xxxx([Pq]xx)xxxxxC_4[ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds;

the trefoil monomer domain comprises the following sequence: C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C_3xx[krln][Gnk]C_4C_5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy];

the thyroglobulin monomer domain comprises the following sequence: C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC_3x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and

wherein α is selected from the group consisting of: w, y, f, and l; ϕ is selected from the group consisting of: d, e, and n; and "x" is selected from any amino acid..

5. The method of claim 1, wherein

the thrombospondin monomer comprises the following sequence: C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6;

the trefoil monomer domain comprises the following sequence: C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C_6; and

the thyroglobulin monomer comprises the following sequence: C_1[qerl]xxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6.

6. The method of claim 1, further comprising linking the identified monomer domains to a second monomer domain to form a library of multimers, each multimer comprising at least two monomer domains;

screening the library of multimers for the ability to bind to the first target molecule; and

identifying a multimer that binds to the first target molecule.

7. The method of claim 6, wherein each monomer domain of the selected multimer binds to the same target molecule.

8. The method of claim 6, wherein the selected multimer comprises three monomer domains.

9. The method of claim 6, wherein the selected multimer comprises four monomer domains.

10. The method of claim 1, further comprising a step of mutating at least one monomer domain, thereby providing a library comprising mutated monomer domains.

11. The method of claim 10, wherein the mutating step comprises recombining a plurality of polynucleotide fragments of at least one polynucleotide encoding a polypeptide domain.

12. The method of claim 1, further comprising,

screening the library of monomer domains for affinity to a second target molecule;

identifying a monomer domain that binds to a second target molecule;

linking at least one monomer domain with affinity for the first target molecule with at least one monomer domain with affinity for the second target molecule, thereby forming a multimer with affinity for the first and the second target molecule.

13. The method of claim 1, wherein the library of monomer domains is expressed as a phage display, ribosome display or cell surface display.

14. The method of claim 1, wherein the library of monomer domains is presented on a microarray.

15. A protein, comprising a non-naturally occurring monomer domain that specifically binds to a target molecule

wherein the target molecule is not bound by a naturally-occurring monomer domain at least 90% identical to the non-naturally occurring monomer domain,

wherein the non-naturally occurring monomer domain is selected from the group consisting of: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain.

16. The protein of claim 15, wherein the monomer domain comprises at least one disulfide bond.

17. The protein of claim 15, wherein the monomer domain comprises at least three disulfide bonds.

18. The protein of claim 15, wherein the monomer domain is 30-100 amino acids in length.

19. The protein of claim 15,

wherein the thrombospondin monomer domain comprises the following sequence:

(wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6;

the trefoil monomer domain comprises the following sequence:

C_1(xx)xxxpxxRxnC_2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and

the thyroglobulin monomer domain comprises the following sequence:

C_1xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vd xx(x)GxxxxGxxxxxgxx(xx)xC_6;

wherein "x" is any amino acid.

20. The protein of claim 19, wherein

C_1-C_5, C_2-C_6 and C_3-C_4 of the thrombospondin monomer domain form disulfide bonds; and

C_1-C_2, C_3-C_4 and C_5-C_6 of the thyroglobulin monomer domain form disulfide bonds.

21. The protein of claim 15,

wherein the thrombospondin monomer domain comprises the following sequence:

(WxxWxx)C_1[Stnd][Vkaq][Tspl]C_2xx[Gq]xx(x)x[Re]x[Rktvm]xC_3[vldr]xxxx ([Pq]xx)xxxxxC_4[ldae]xxxxxx(x)xxxC_5(x)xxxxC_6, wherein C_1-C_5, C_2-C_6 and C_3-C_4 form disulfide bonds;

the trefoil monomer domain comprises the following sequence:

C_1(xx)xxx[Pvae]xxRx[ndpm]C_2[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C_3xx[krln][G nk]C_4C_5[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C_6[Fy];

the thyroglobulin monomer domain comprises the following sequence: C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[αhp]xPxC_2xxxGx[α]xx[vkrl]QC_3x(x[sa]xxx)xx[gas]xC_4[α]C_5V[Dnα]xx(x)Gxxxx[ϕg]xxxxxgxx(xx)xC_6, wherein C_1-C_2, C_3-C_4 and C_5-C_6 form disulfide bonds; and

wherein α is selected from the group consisting of: w, y, f, and l; ϕ is selected from the group consisting of: d, e, and n; and "x" is selected from any amino acid.

22. The protein of claim 15,

wherein the thrombospondin monomer comprises the following sequence: C_1[nst][aegiklqrstv][adenpqrst]C_2[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]xC_3xxxxxxxxx(xxxxxxx)C_4xxxxxxxxx(xx)C_5xxxxC_6;

the trefoil monomer domain comprises the following sequence: C_1([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][fiklqrtv][dnpqs]C_2[agiy][flpsvy][dknpqs][adfghlp][aipv][st][aegkpqrs][adegkpqs][deiknqt]C_3[adefknqrt][adegknqs][gn]C_4C_5[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afinprstv][degklns][afiqstv][iknpv]w)C_6; and

the thyroglobulin monomer comprises the following sequence: C_1[qerl]xxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxx[Yfhp]xPxC_2xxxGx[Yf]xx[vkrl]QC_3x(x[sa]xxx)xx[Gsa]xC_4[Wyf]C_5V[Dnyfl]xx(x)Gxxxx[Gdne]xxxxxgxx(xx)xC_6.

23. An isolated polynucleotide encoding the protein of claim 15.

24. A library of proteins comprising non-naturally-occurring monomer domains, wherein the monomer domain is selected from the group consisting of: a thrombospondin monomer domain, a trefoil monomer domain, and a thyroglobulin monomer domain,

wherein the thrombospondin monomer domain comprises the following sequence: (wxxWxx)C_1sxtC_2xxGxx(x)xRxrxC_3xxxx(Pxx)xxxxxC_4xxxxxx(x)xxxC_5(x)xxxxC_6;

the trefoil monomer domain comprises the following sequence: C_1(xx)xxxpxxRxnC2gx(x)pxitxxxC_3xxxgC_4C_5fdxxx(x)xxxpwC_6f; and

the thyroglobulin monomer domain comprises the following sequence: C_1xxxxxxxxxxxxxxxx(xxxxxxxxxx)xxxxxxxyxPxC_2xxxGxyxxxQC_3x(x)s(xxx)xxgxC_4WC_5Vdxx(x)GxxxxGxxxxxgxx(xx)xC_6;

wherein "x" is any amino acid.

25. The library of claim 24, wherein each monomer domain of the multimers is a non-naturally occurring monomer domain.

26. The library of claim 24, wherein the library comprises a plurality of multimers, wherein the multimers comprise at least two monomer domains linked by a linker.

27. The library of claim 24, wherein the library comprises at least 100 different proteins comprising different monomer domains.

28. A library of polynucleotides encoding the library of proteins of claim 24.

Figure 1

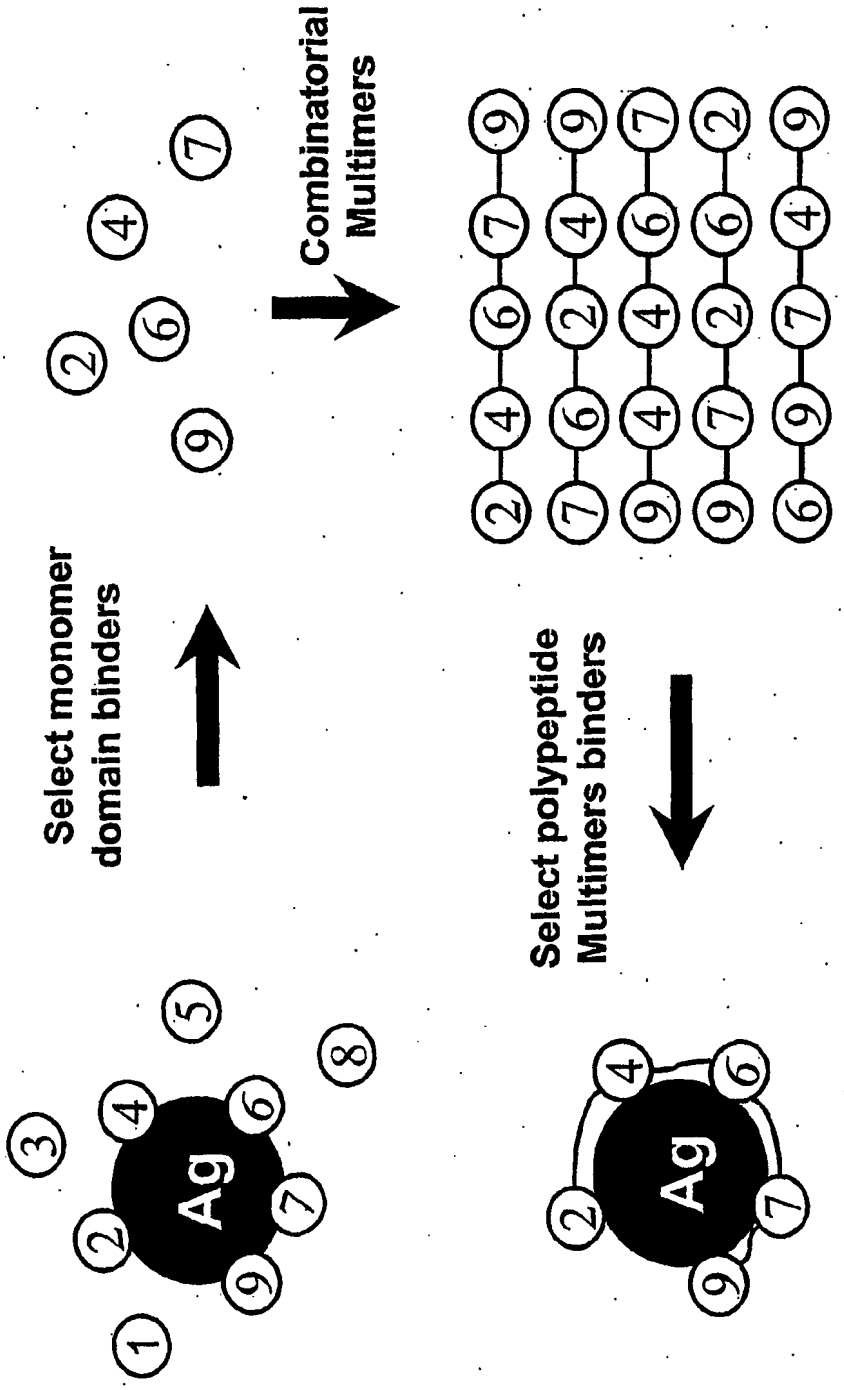

Combinatorial Multimers
Select monomer domain binders
Select polypeptide Multimers binders
Ag
Ag

Figure 2

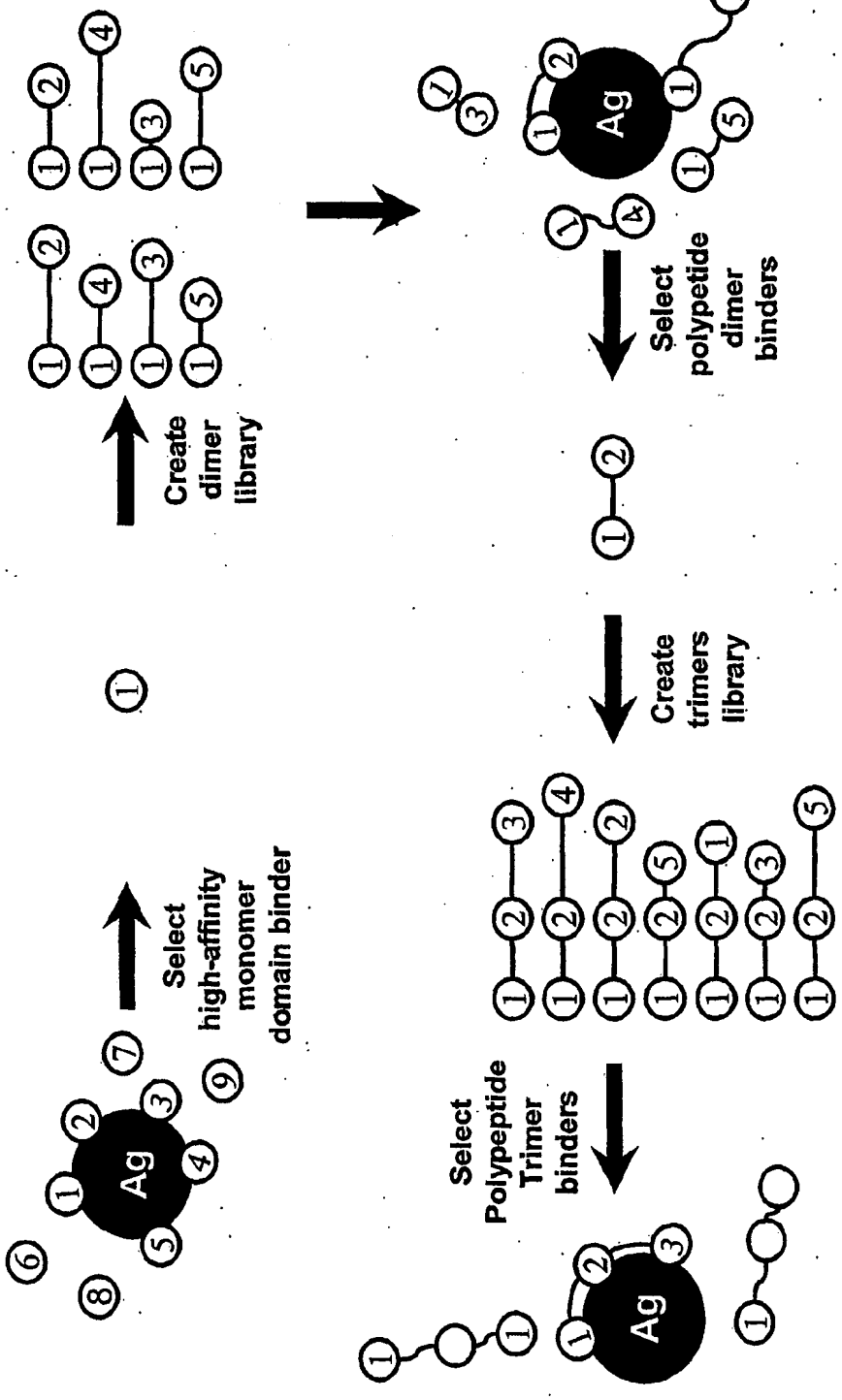

Ag
Select high-affinity monomer domain binder
Create dimer library
Select polypetide dimer binders
Create trimers library
Select Polypeptide Trimer binders
Ag
Ag

Figure 3

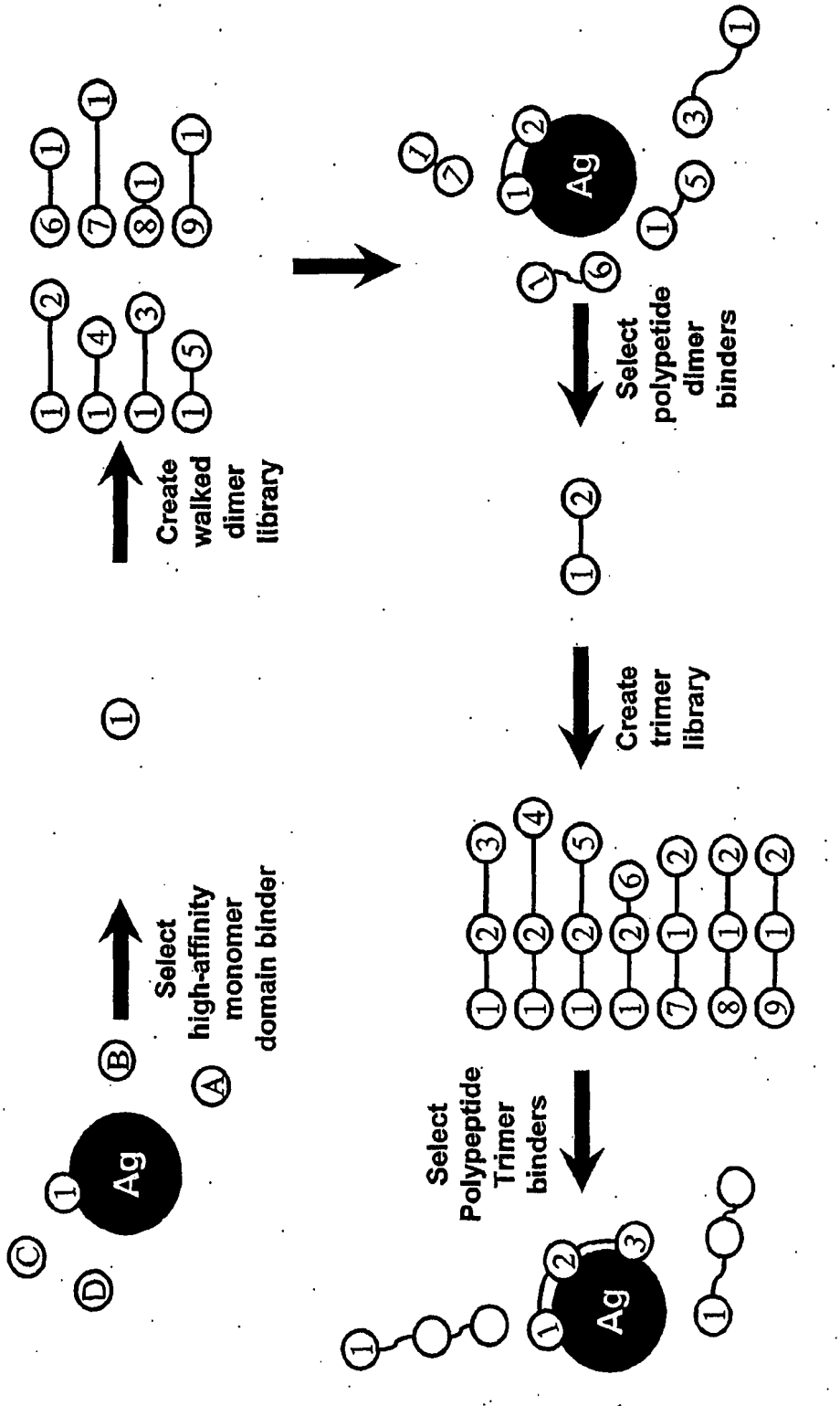

Ag
A
B
C
D
Select high-affinity monomer domain binder
Create walked dimer library
Select polypetide dimer binders
Create trimer library
Select Polypeptide Trimer binders

Figure 4

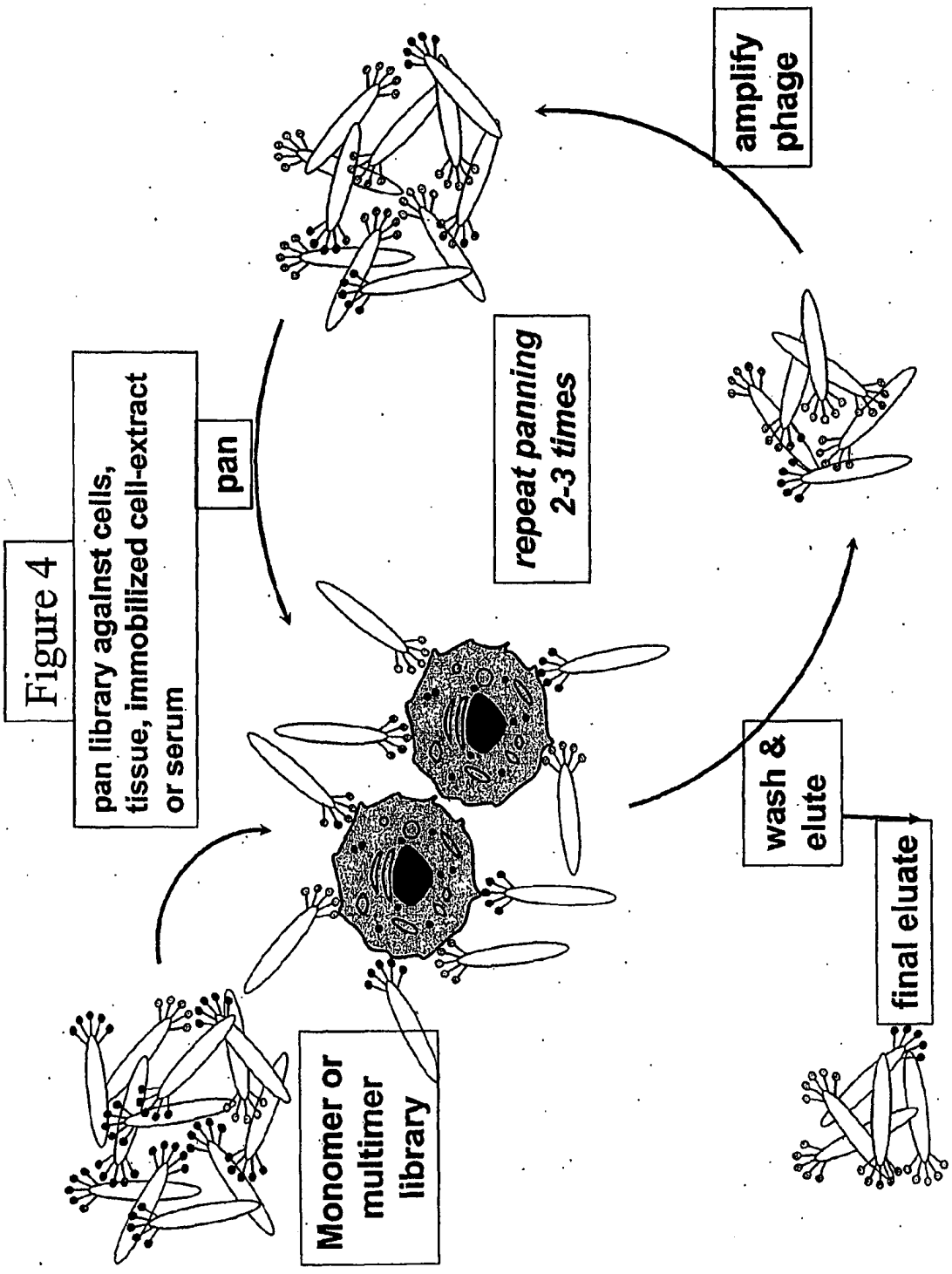

pan library against cells, tissue, immobilized cell-extract or serum
pan
amplify phage
repeat panning 2-3 times
wash & elute
final eluate
Monomer or multimer library

Figure 5

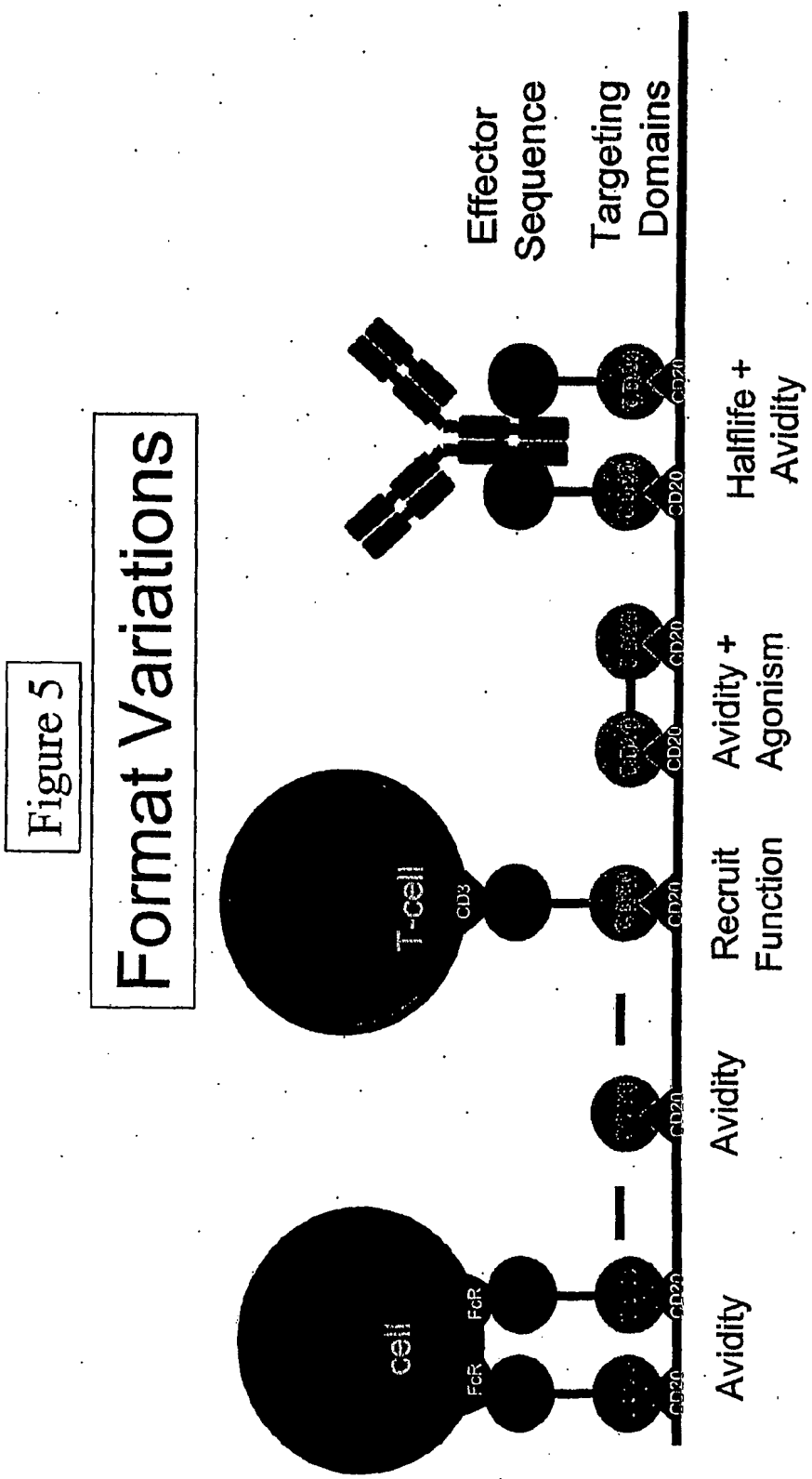

Format Variations
Effector Sequence
Targeting Domains
T-cell
CD3
Cell
FcR
FcR
CD20
Avidity
Avidity
Recruit Function
Avidity + Agonism
Halflife + Avidity

Figure 6

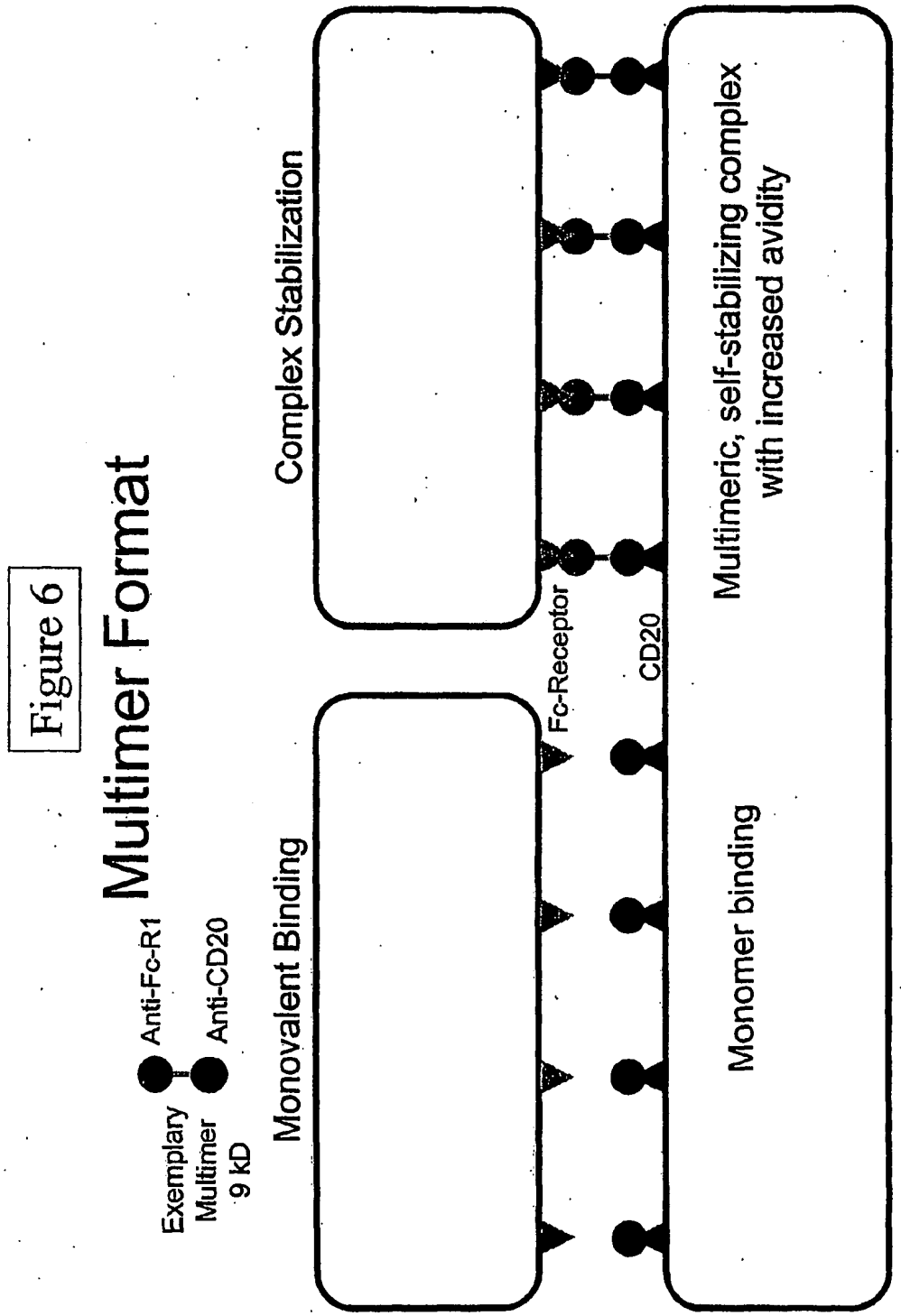

Multimer Format
Anti-Fc-R1
Anti-CD20
Exemplary Multimer 9 kD
Monovalent Binding
Complex Stabilization
Fc-Receptor
CD20
Monomer binding
Multimeric, self-stabilizing complex with increased avidity

Figure 7

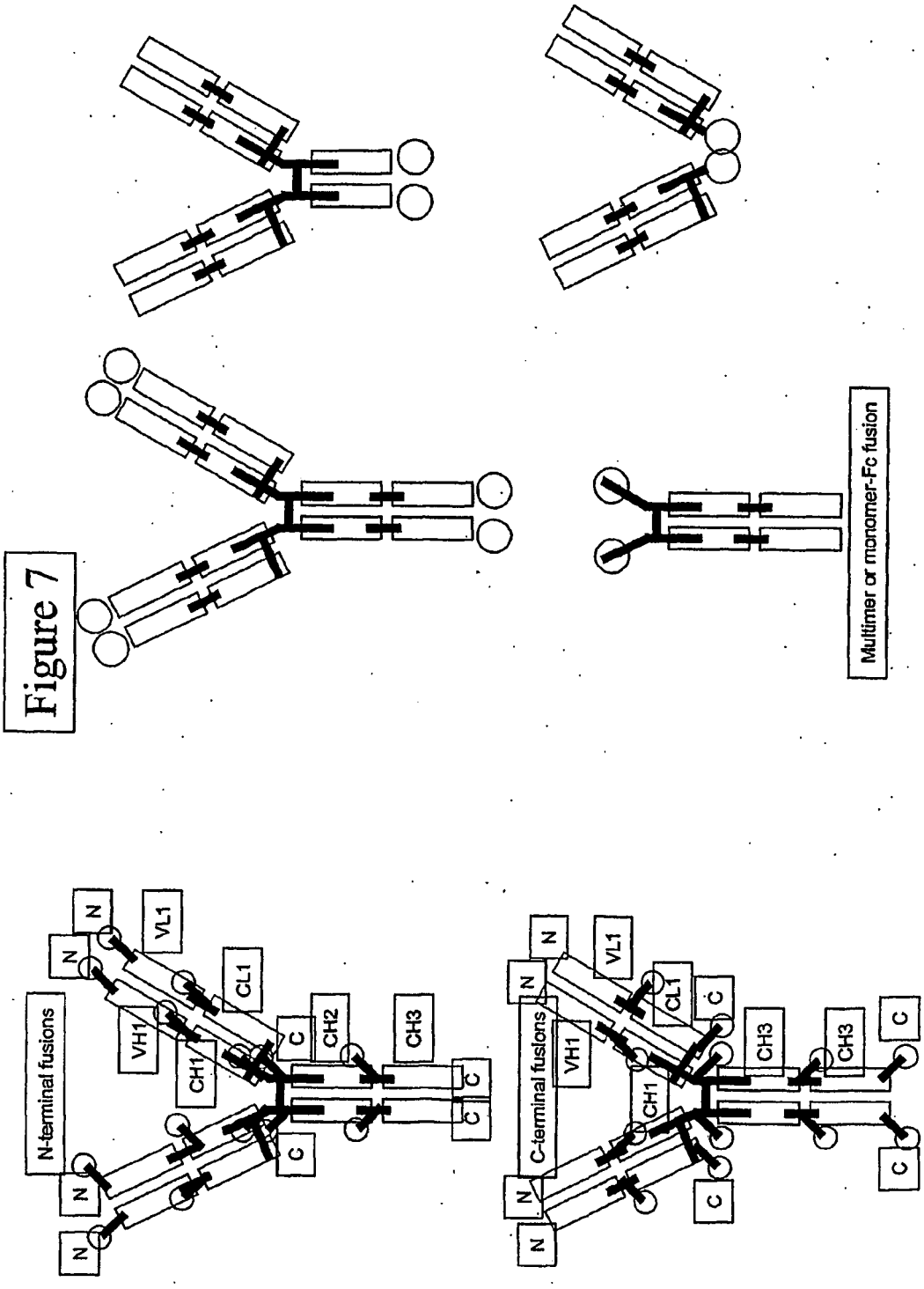

Multimer or monomer-Fc fusion
N-terminal fusions
C-terminal fusions
N
VL1
CL1
VH1
CH1
CH2
CH3
C

Figure 8

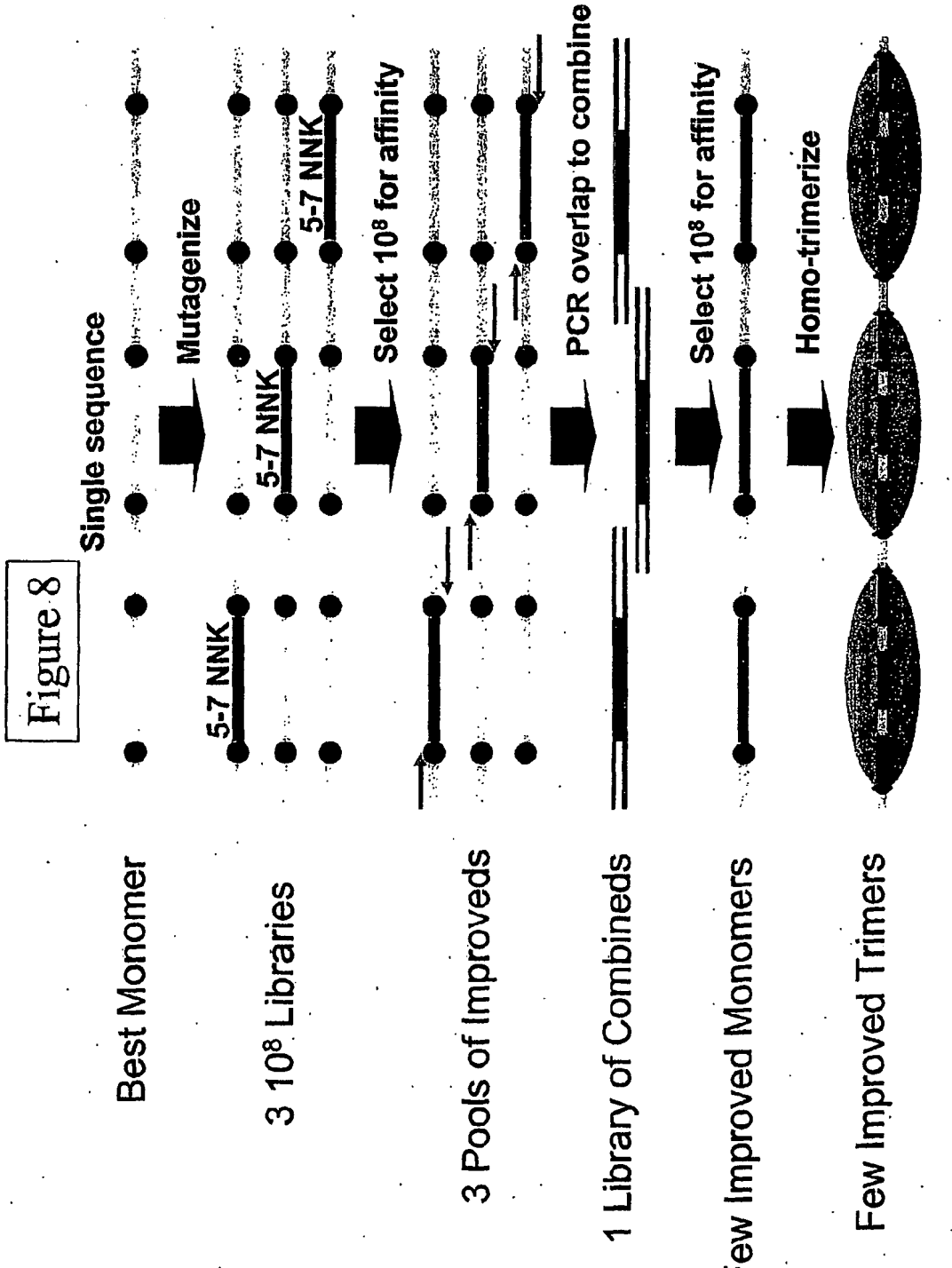

Single sequence
Mutagenize
5-7 NNK
5-7 NNK
5-7 NNK
Select 10^8 for affinity
PCR overlap to combine
Select 10^8 for affinity
Homo-trimerize
Best Monomer
3 10^8 Libraries
3 Pools of Improveds
1 Library of Combineds
Few Improved Monomers
Few Improved Trimers

Figure 9

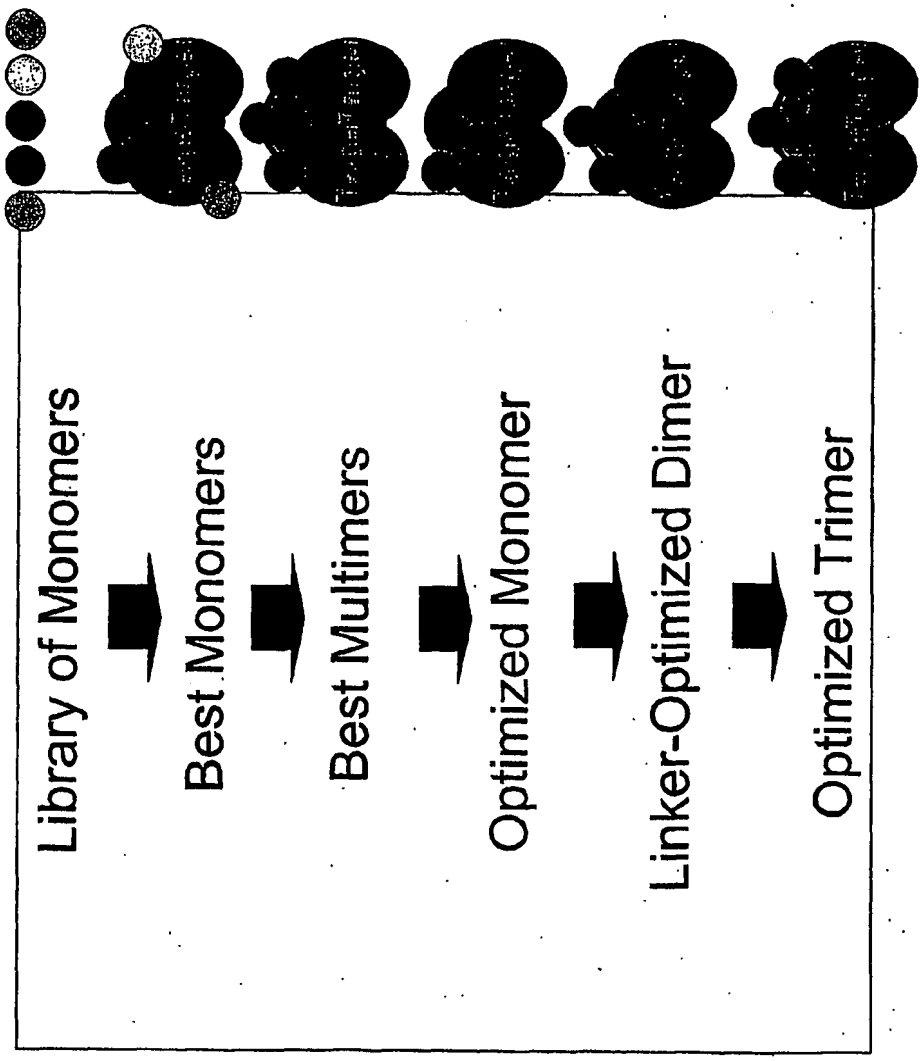

Library of Monomers
Best Monomers
Best Multimers
Optimized Monomer
Linker-Optimized Dimer
Optimized Trimer

Figure 10

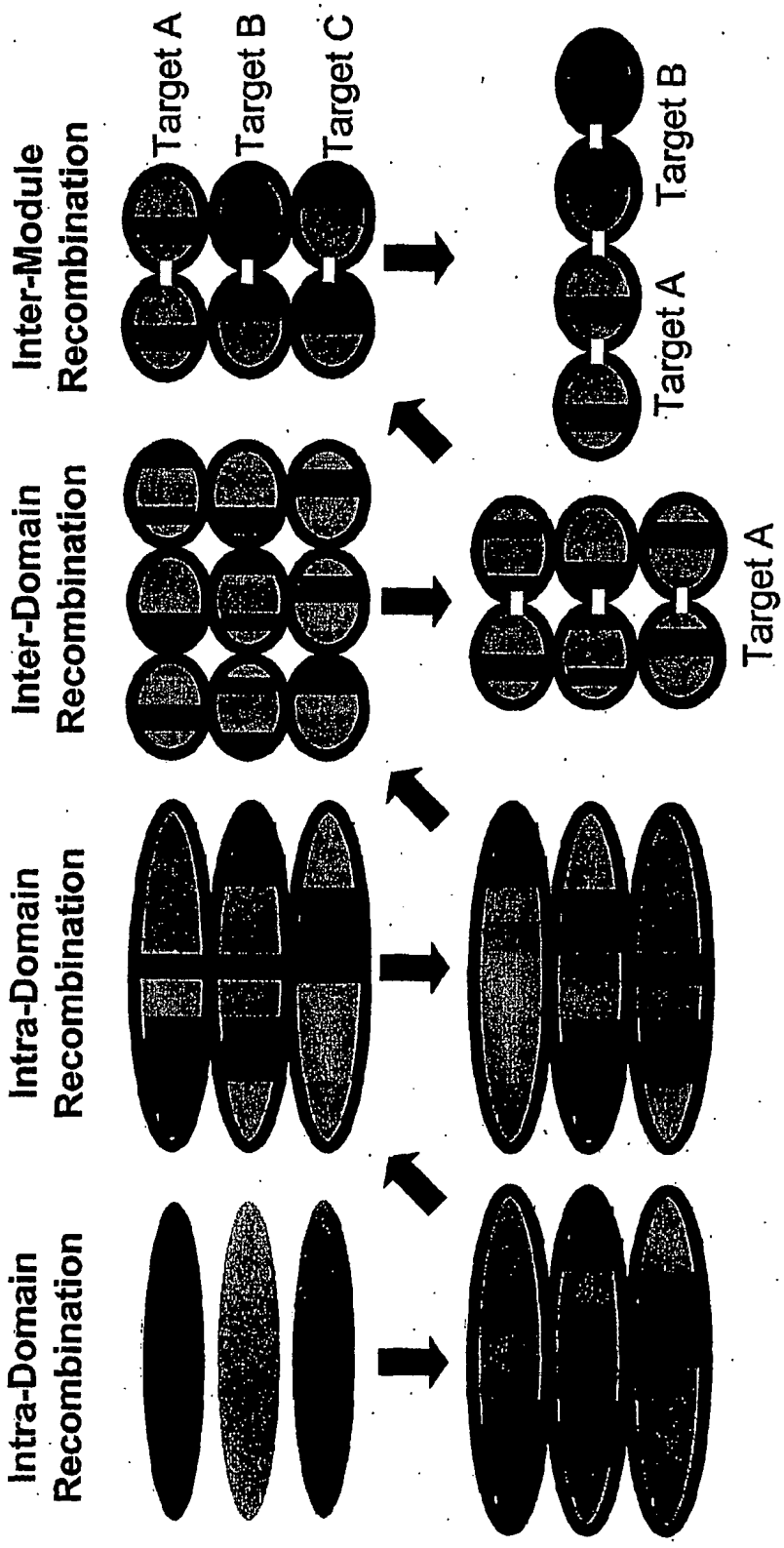
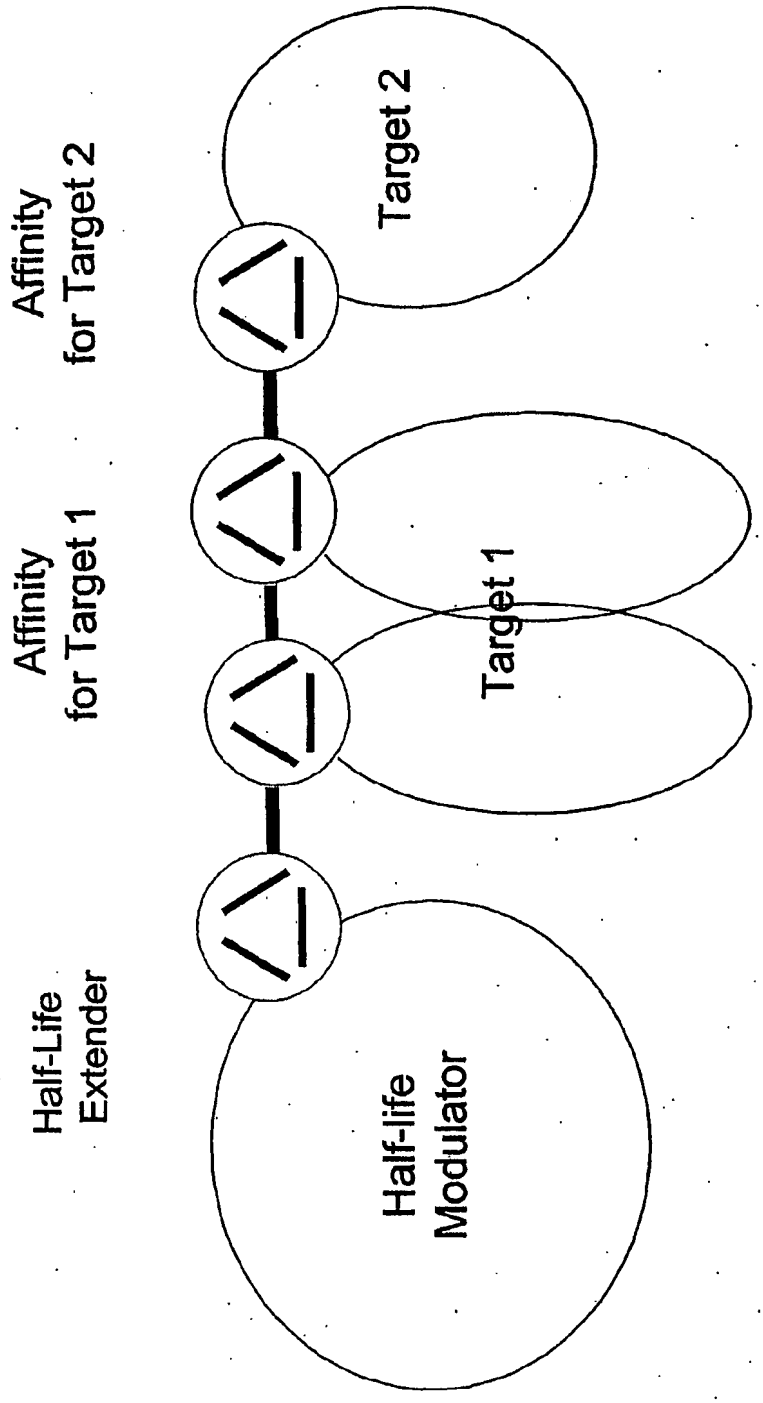

Intra-Domain Recombination
Intra-Domain Recombination
Inter-Domain Recombination
Inter-Module Recombination
Target A
Target B
Target C
Target A
Target B
Target A

Figure 11


Half-Life Extender
Affinity for Target 1
Affinity for Target 2
Half-life Modulator
Target 1
Target 2